Accept 3/23/89

AD-A207 937

REPORT DOCUMENTATION PAGE

1b. RESTRICTIVE MARKINGS: None	
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION / AVAILABILITY OF REPORT: Unlimited (2)
2b. DECLASSIFICATION / DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S): NONE	5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 89-0615
6a. NAME OF PERFORMING ORGANIZATION: Worcester Polytechnic Institute	6b. OFFICE SYMBOL (If applicable):
7a. NAME OF MONITORING ORGANIZATION: Dr. Abraham Waksman/NM (202)767-5027	
6c. ADDRESS (City, State, and ZIP Code): Worcester Polytechnic Institute, EE Dept. 100 Institute Road Worcester, Massachusetts 01609	7b. ADDRESS (City, State, and ZIP Code): AFOSR/NM, Building 410 Bolling AFB, DC 20332-6448
8a. NAME OF FUNDING / SPONSORING ORGANIZATION: AFOSR	8b. OFFICE SYMBOL (If applicable): AFOSR/NM
9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: AFOSR-89-0037	
8c. ADDRESS (City, State, and ZIP Code): AFOSR/NM, Building 410 Bolling AFB, DC 20332-6448	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. 61102F; PROJECT NO. 2304; TASK NO. ~~2304~~ A7; WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)

Application of Multi-Channel Hough Transform to Stereo Vision

12. PERSONAL AUTHOR(S)

Nasrabadi, Nasser Mohammadi

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 11/1/88 TO 10/30/89	1989, March 01	102

16. SUPPLEMENTARY NOTATION

NONE

17. COSATI CODES		
FIELD	GROUP	SUB-GROUP

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)

Stereo Vision, Computer Vision, Robotics, Stereo-Matching, Relaxation, Hough Transform, Binocular Stereo.

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

A major issue in any stereo vision system is the correspondence problem. In this report a feature-based stereo vision technique is described where curve-segments are used as the feature primitives in the matching process. The local characteristics of the curve-segments are extracted by the Generalized Hough Transform (R-table) representation of the curve-segment. The left image and the right image are first filtered by using several Laplacian of a Gaussian operator ($\nabla^2 G$) of different widths (channels). Curve-segments are extracted by a tracking algorithm and their centroids are obtained. At each channel, the Generalized Hough Transform of each curve-segment in the left and the right image is evaluated. This is done by calculating the R-table representation of each curve-segment using the centroid of the curve-segment as the reference point. The R-table, is used as a local feature vector in representing the distinctive characteristics of the curve-segment. Initial node assignments are formed between the left curve-segments and the right curve-segments if they satisfy the epipolar constraint and their R-tables satisfy a similarity measure. The epipolar constraint on the centroids of the curve-segment and the channel size is used to limit the searching space in the right image.

To resolve the ambiguity of the false targets (multiple matches) a relaxation technique is used where the initial scores of the node assignments are updated by the compatibility measures between the centroids of the curve-segments. The node assignments with the highest score are chosen as the matching curve-segments. This algorithm is believed to be an improvement of the Marr-Poggio-Grimson algorithm.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT	21. ABSTRACT SECURITY CLASSIFICATION
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT ☐ OTIC USERS	Unclassified
22a. NAME OF RESPONSIBLE INDIVIDUAL: Dr. Abraham Waksman	22b. TELEPHONE (Include Area Code): (202) 767-5027 / 22c. OFFICE SYMBOL: nm

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete. SECURITY CLASSIFICATION OF THIS PAGE

89 5 15 216

AFOSR-TR- 89-0615

APPLICATION OF MULTI-CHANNEL HOUGH TRANSFORM TO STEREO VISION

Dr. Nasser M. Nasrabadi

Computer Vision Research Group
Department of Electrical Engineering
100 Institute Road
Worcester Polytechnic Institute
Worcester, MA 01609
Tel. (508) 831 5257

March 1989

Abstract

A major issue in any stereo vision system is the correspondence problem. In this report a feature-based stereo vision technique is described where curve-segments are used as the feature primitives in the matching process. The local characteristics of the curve-segments are extracted by the Generalized Hough Transform (R-table) representation of the curve-segment. The left image and the right image are first filtered by using several Laplacian of a Gaussian operator ($\nabla^2 G$) of different widths (channels). Curve-segments are extracted by a tracking algorithm and their centroids are obtained. At each channel, the Generalized Hough Transform of each curve-segment in the left and the right image is evaluated. This is done by calculating the R-table representation of each curve-segment using the centroid of the curve-segment as the reference point. The R-table, is used as a local feature vector in representing the distinctive characteristics of the curve-segment. Initial node assignments are formed between the left curve-segments and the right curve-segments if they satisfy the epipolar constraint and their R-tables satisfy a similarity measure. The epipolar constraint on the centroids of the curve-segment and the channel size is used to limit the searching space in the right image.

To resolve the ambiguity of the false targets (multiple matches) a relaxation technique is used where the initial scores of the node assignments are updated by the compatibility measures between the centroids of the curve-segments. The node assignments with the highest score are chosen as the matching curve-segments. This algorithm is believed to be an improvement of the Marr-Poggio-Grimson algorithm.

Accepted
3/23/89

APPLICATION OF MULTI-CHANNEL HOUGH TRANSFORM TO STEREO VISION

Dr. Nasser M. Nasrabadi

A major issue in any stereo vision system is the correspondence problem. In this report a feature-based stereo vision technique is described where curve-segments are used as the feature primitives in the matching process. The local characteristics of the curve-segments are extracted by the Generalized Hough Transform (R-table) representation of the curve-segment. The left image and the right image are first filtered by using several Laplacian of a Gaussian operator ($\nabla^2 G$) of different widths (channels). Curve-segments are extracted by a tracking algorithm and their centroids are obtained. At each channel, the Generalized Hough Transform of each curve-segment in the left and the right image is evaluated. This is done by calculating the R-table representation of each curve-segment using the centroid of the curve-segment as the reference point. The R-table, is used as a local feature vector in representing the distinctive characteristics of the curve-segment. Initial node assignments are formed between the left curve-segments and the right curve-segments if they satisfy the epipolar constraint and their R-tables satisfy a similarity measure. The epipolar constraint on the centroids of the curve-segment and the channel size is used to limit the searching space in the right image.

To resolve the ambiguity of the false targets (multiple matches) a relaxation technique is used where the initial scores of the node assignments are updated by the compatibility measures between the centroids of the curve-segments. The node assignments with the highest score are chosen as the matching curve-segments. This algorithm is believed to be an improvement of the Marr-Poggio-Grimson algorithm.

1. INTRODUCTION

In applications, such as robotics and automation, three dimensional information about the environment is essential for the movement of robots and object inspection. Depth information is important for the control of the robot arm as well as for object modeling and recognition. The research proposed in this report concerns the development of algorithms for obtaining the distance from the camera to the objects in the scene using a pair of stereo images. The absolute or the relative depth information can be obtained from vision techniques like monocular cues, motion parallex, stereo vision, structure light, and laser range finders.

Monocular cues, for example structural texture [1], shading [2], shadow [3] and line drawings [4], are ill-posed problems [41]-[43]. Solutions to these problems are very difficult to obtain. In addition, these cues appear only in some images. Structure light [5] and laser range finders [6] are very effective in obtaining the depth, but their use is also limited, to some particular environments. However, motion parallex [9] and stereo vision are the most passive way, to acquire depth information. In this report, we are

only interested in stereo vision algorithms.

The major problem in stereo computation is to find the corresponding points in the stereo images. The corresponding points are the projections of a single point in the three-dimensional scene. The difference in the positions of the two corresponding points in their respective images is called disparity. Disparity is a function of both the position of the point in the scene and of the position, orientation, and physical characteristics of the stereo cameras. The research proposed in this report concerns the development of algorithms to solve the correspondence problem in a stereo vision system.

Investigation into stereo vision algorithms is a significant research problem. There has been a number of applications that use stereo vision systems, for example the three dimensional inspection of VLSI chips by using a pair of Scanning Electron Microscope (SEM) stereo images. Data obtained from a stereo system can be used in conjunction with a robot arm to perform object manipulations. Recognition of objects in 3-D when they are overlapping or touching is another research problem. Another application for stereo vision systems is to create 3-D models of the work cell for robot arm trajectory planning.

Stereo techniques have a number of applications in aerial photogrammetry [66] (see Appendix IV). For example photogrammetrists use stereoplotters to obtain the surface topography of the environment. These stereo plotters are operated manually and 3-D information can only be obtained at a few interesting feature points where correspondence is solved by the operator. There is a great need for an automated stereoplotter and the solution is a stereo vision algorithm. In X-ray photogrammetry [67], stereo X-ray images are used to obtain the location of foreign objects, such as bullets in the body.

The camera's geometry [7]-[8] in any stereo vision system is very important because it is possible to constrain the search for matching pairs of corresponding image points to one dimension. In this paper, it is assumed that the two cameras are mounted such that their focal axes are parallel and the distance between the two cameras (baseline), b, is fixed as shown in Figure 1. This is known as the parallel axis

geometry. A detailed description of the camera's calibration is given in Appendix I.

Any point in the three dimensional world space, together with the centers of projection of the two camera systems, defines a plane called an epipolar line. In the parallel axis geometry the epipolar lines are parallel to the scan lines. Thus the search for finding corresponding points is unidirectional as shown in Figure 1.

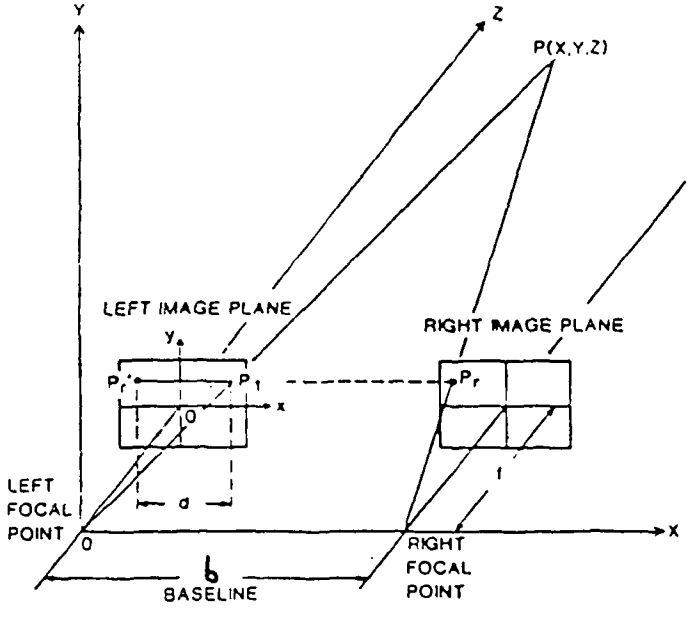


Fig. 1. **Parallel axes method**, the cameras are set up such that their focal axes are parallel and the line joining the focal centers is perpendicular to it [22].

Consider the point $P(X,Y,Z)$ in the world coordinate system that is imaged into point P_r and P_l in the right and left image coordinate plane respectively. The distance $P'_r P_l$, where P'_r is the transformed location of P_r in the left image plane, is known as disparity. It can easily be shown [7]-[8] that the distance Z is inversely proportional to the disparity. This is shown by expression

$$Z = b \frac{f}{P'_r P_l} \tag{1}$$

where b is the distance between the origin of the cameras. Thus the points which are nearer to the camera will have a larger disparity than the points which are farther away from the camera. Therefore once the location of the same target points P_l and P_r are identified, the distance Z can be calculated.

In this report a high level stereo technique or structural matching technique is proposed. In our proposed method, a global match between the left and the right image curve-segments is achieved by using a graph matching technique [47]-[56]. The centroids of the curve-segments in the left image form a graph that represents the geometrical relationships of the left curve-segments, and portrays the structural information about the object. A similar graph is formed for the right image. Sub-graph isomorphisms are then obtained by using a clique finding technique [47] or a relaxation technique [48].

The proposed technique is believed to perform better than line-based stereo techniques because the iconic properties of each zero-crossing pixel are stored in the Generalized Hough representation of the curve-segment so no information is lost. Also the ambiguity in identifying continuous long curve-segments in a pair of stereo images should be much less than that in identifying short line-segments.

In this report, a review of the previous work on stereo vision algorithms is given in Section 2. The proposed research is discussed in Section 3, and a detailed study is given in the subsequent subsections. In Section 4, experimental results are presented. Finally in section 5, conclusion is given.

2. A REVIEW OF STEREO VISION TECHNIQUES

To solve the correspondence problem, one can divide the algorithms into the following categories:

[1] Cooperative algorithms [9] - [12].

[2] Area-based matching [13] - [15].

[3] Feature-based matching [16] - [23].

[4] Gradient techniques [24].

[5] Others [26], [32] - [33].

The human stereo vision system has been an interest to psychophysicists for a number of decades [9]-[12]. The human visual system, (HVS), has several powerful depth sensing cues. These cues can be classified into Monocular and Binocular cues. One such binocular cue is stereopsis. This is based on the geometrical fact that two-dimensional projections of a three-dimensional object on the left and the right retina differ in their horizontal positions. This horizontal shift between corresponding points in the two retinal images is called retinal disparity. Other binocular depth cues are the complex vergence control of the eye and correlative accommodation, (differential focusing of two eyes) [10 pp.144], but these cues are not as powerful as stereopsis.

Monocular cues, such as gradient texture [9], perspective cues, shape from shading, shadow, occlusion, line-drawings, and movement parallax are very important in depth perception. Monocular movement parallax is a particularly strong depth cue. Its action is very similar to stereopsis. The movement of the eye's position creates motion disparities such that objects closer to the eye appear to have moved faster than objects that are further away. These motion disparities will produce a depth sensation in the human visual system.

The binocular cues and monocular cues interact with each other in unknown complex ways to produce the relative depth sensation. Julesz [10] set up an experiment where the monocular depth cues were removed by using his computer generated random-dot stereograms. Figure 2 shows a pair of random-dot stereogram images. These images have identical random-dot textures. Certain areas of these textures are identical and shifted relative to each other in the horizontal direction as though they were solid sheets. These stereograms when viewed monocularly appear as random-dots, but when the pair is seen through a stereoscope or crossing the eyes, a floating square in space above the plane of the background will be perceived [10 pp.156]. This proves that binocular disparity alone can cause sensation of depth. There are also no monocular structures in the stereogram that can be matched by vergence control. Julesz also performed experiments to investigate the range over which one can fuse two images, the expansion and rotation of one stereogram with respect to the other, and the fusion of the stereogram if it contains the same frequency components [10 pp.98].

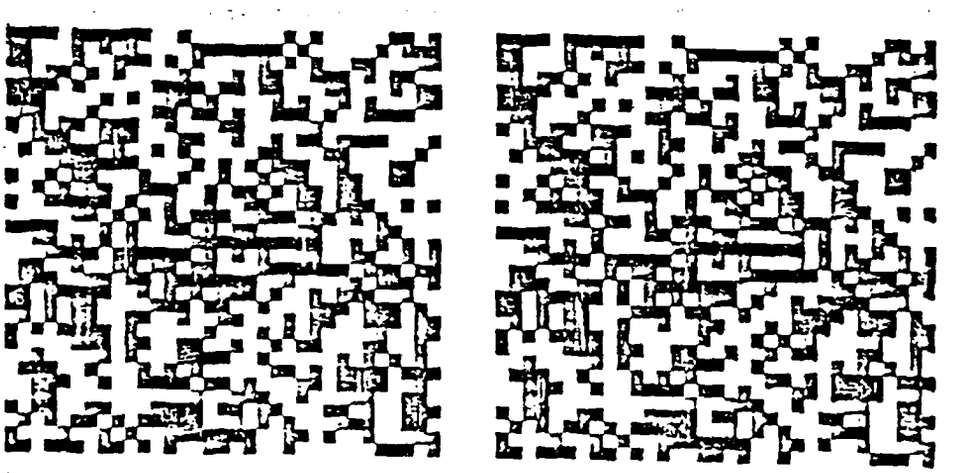


Fig. 2. Random-dot stereogram of 33×33 resolution, the center square is only 15×15 and displaced by a few pixels [10 PP. 156].

Julesz proposed a magnetic dipole model where tiny compass needles were imagined to be suspended at their centers so that the needles can rotate in any direction in or out from a plane. Two networks of magnetic dipoles were arranged. One for the left image and another for the right image. The two networks were overlayed and the polarity of each needle, (North or South), was chosen according to the intensity of the image, (black or white), at that location. The end points of neighboring needles on each side were coupled together by springs to produce a global fusion or to satisfy the continuity rule [10]. His model works by some random shift aligning of certain similar dipole arrays, which are said to be interlocked. Searches are then made for other similar dipole arrays by performing horizontal shifts. The above cooperative model will not be very useful in our robotics application because the objects that are viewed in the robotics application have a lot of monocular cues that could be used for stereo fusion.

Marr and Poggio [11] proposed a cooperative algorithm where a parallel and interconnected network of processors were used to fuse a pair of binary stereograms. In their algorithm, they introduced and implemented three rules [11 pp.115] as given below:-

1) Compatibility: Black dots can match only black dots.

2) Uniqueness: Almost always a black dot from one image can match no more than one black dot from the other image.

3) Continuity: The disparity of the matches varies smoothly almost everywhere over the image.

In Figure 3, continuous vertical and horizontal lines represent lines of sight from the left and the right eye. The intersections of these lines correspond to possible disparity values. The dotted diagonal lines are lines of constant disparity. At each intersection, or node, a processor is placed such that all the processors at the nodes along each vertical or horizontal line will inhibit each other, and connections along the dotted lines in Figure 3 will exhibit each other. This network of processors is left to run iteratively by first initializing it, (putting a 1 wherever two black dots match and 0 at all other places). Each processor adds up the 1's in its excitatory neighborhood, adds up the 1's in its inhibitory neighborhoods, and subtracts the resulting figures as represented by

the iterative relation,

$$C^{t+1}_{x,y;d} = \sigma \left\{ \sum_{x',y';d' \in S(x,y;d)} C^{t}_{x',y';d'} - \epsilon \sum_{x',y';d' \in O(x,y;d)} C^{t}_{x',y';d'} + C^{0}_{x,y;d} \right\} \qquad (2)$$

where $C^{t}_{x,y;d}$ denotes the state of the call corresponding to position (x,y), disparity d, and tth iteration. $S(x,y,d)$ and $O(x,y,d)$ are the local excitatory and inhibitory neighborhoods respectively. C^0 represents the initial states, ϵ is an inhibition constant, and σ is a threshold function.

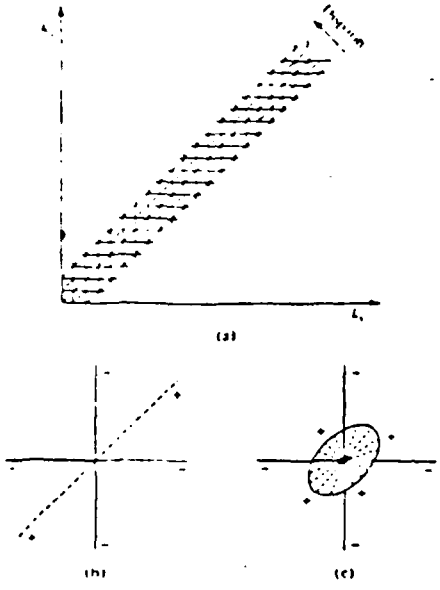


Fig. 3. The continuous vertical and horizontal lines represent lines of sight from the left and the right eye. The intersections of these lines corespond to possible disparity values. The dotted diagonal lines are lines of constant disparity [11].

Marr-Poggio tested their algorithm on a pair of stereograms. The technique was successful to fuse the stereo images after 14 iterations. The algorithm performs better on the high density stereograms than on sparse stereograms, but it becomes very slow as the number of nodes increases. The two cooperative algorithms mentioned above are very successful in simulating binocular cues of the eye, but there are several other cues that are important which could be used to improve the stereo vision algorithms.

The reason for proposing such a point-to-point cooperative matching technique was that the human binocular vision system can easily fuse random dot stereograms. Since these stereograms do not show any structural information before fusion, it is suggested that the stereopsis is performed at a very low level such as point-to-point matching of gray-level intensity of the image. However, Hubel and Wiesel [68]-[69] have demonstrate that there are several different neuron cells (known as simple, complex and hypercomplex cells) in the primary visual cortex that respond best to a specific directional stimuli, such as vertical or horizontal oriented light bars (see Appendix III).

Barnard [62]-[63] presented a stochastic optimization technique (simulated annealing) to fuse a pair of stereo images. It is now known [75] that a Markov Random field (MRF) defined over a neighborhood is equivalent to the Gibbs distribution of the whole system. There are stochastic techniques such as the simulated annealing [74] where global minimization can be obtained proveded the system is a Gibbsian. Thus, a functional (energy) similar to regularization theory introduced by Poggio [41]-[43] is minimized whose solution is the stereo correspondence. The functional energy is given by

$$E = \sum_{i,j} \| I_l(i,j) - I_r(i,j+D(k)) \| + \lambda \| \nabla D(k) \| \qquad (3)$$

the first term in the summation represents the photogrammetric constraint and the second term smoothness constraint.

Prazdny [61] introduced a parallel stereo algorithm allowing the disparities from the neighboring point to give support if their disparities were similar. No inhibitory score was allowed because if disparities are not similar they belong to a different physical object. He introduced the coherence constraint which states that the neighboring disparities of elements corresponding to the same 3-D object must be similar. In his

algorithm first he finds all potential disparities for each feature point (edges) in the left image. Associated with each possible disparity is an activity cell whose value indicates the amount of support the particular disparity receives from its neighbors.

The local support for each point i with a given disparity d_i is calculated by

$$S(i,j) = \frac{1}{c\,|i-j|\sqrt{2\pi}}\; e^{-\frac{|d_j - d_i|^2}{2c^2|i-j|^2}} \tag{4}$$

where j is a neighboring point. For any feature point in the left image the maximum local support from all the neighboring points are added together for each possible disparity. After the support for all possible disparities at a given point has been determined the disparity with the largest support (the highest value in the associated activity cell) is chosen as the most likely disparity at that point.

In area-based matching, cross-correlation is used to determine matches between windows in one image with windows in the other. One major disadvantage of the area-based stereo technique is the computation of the cross-correlation at each image sample. To reduce the computation cost, cross correlation is applied only to pixels with high local variance [27], or to edge pixels and their neighbors [28]. Multiresolution cross-correlation or binary correlation can also be used [15]. Area based matching performs well only when the scene is smoothly varying and continuous. But in applications, such as robotics where there are several objects at different depths as well as occlusion, the area correlation does not perform well.

The correlation measures that are commonly used are

$$COR = \sum (X_i * Y_i) \tag{5}$$

which can be normalized by the means of the samples

$$COR = \sum_i (X_i - \bar{X}) * (Y_i - \bar{Y}) \tag{6}$$

or by the second moments of the samples

$$COR = \frac{\sum_i (X_i * Y_i)}{\sqrt{\sum_i X_i^2 * \sum_i Y_i^2}} \tag{7}$$

where X_i and Y_i represent a pair of stereo image with means of $\bar{X}$ and $\bar{Y}$ respectively.

Rather than using correlation measures, difference measure can also be used such as root-mean-square (RMS) error;

$$RMS = \sqrt{(\frac{1}{n} \sum_i (X_i - Y_i)^2} \tag{8}$$

which can also be normalized by the means of the samples.

$$RMS = \sqrt{\frac{1}{n} \sum_i ((X_i - \bar{X}) - (Y_i - \bar{Y}))^2} \tag{9}$$

Absolute difference is also used.

$$AD = \sum_i \frac{(X_i - Y_i)}{n} \tag{10}$$

It too can be normalized by the means

$$AD = \sum_i \frac{((X_i - \bar{X}) - (Y_i - \bar{Y}))}{n} \tag{11}$$

Feature based stereo algorithms are believed to be computationally less costly than area-correlation and have better accuracy because features, such as edges, can be detected to a sub-pixel accuracy. Two feature based methods have become very popular. One is by Marr-Poggio-Grimson [29]-[30], and the other by Baker and Binford [31]. Both of these techniques are edge matching methods. The major difference in the two algorithms is the way in which the edges are matched.

In the Marr-Poggio-Grimson (MPG) matching algorithm, edge pixels that have the same edge polarity, have approximately the same orientation, and lie on the same epipolar line in the left image, are matched with the edge pixels in the right image. The matching process is done in several channels in order to use the coarse-to-fine strategy. In coarse-to-fine strategy, coarse features are first matched, and then the results are used to converge the matching of finer features (see Appendix II).

A major drawback of the MPG algorithm is that the matching process in each channel is performed locally between a zero-crossing pixel in the left image and a zero-crossing pixel from the right image. This was pointed out by Mayhew and Frisby [23],

and they proposed a modified MPG algorithm by including the figural continuity of the zero-crossings. In their algorithm, a zero-crossing pixel in the left image is said to be matched with a zero-crossing pixel in the right image provided that the neighboring zero-crossing pixels lying on the same edge are also matched. They found that as the number of neighboring matching points is increased as a constraint, the number of zero-crossing mismatches is drastically reduced. The proposed algorithm in this paper incorporates figural continuity because our matching primitives are curve-segments. Therefore, the number of false targets or mismatches is expected to be much smaller than that of the MPG algorithm.

Kim and Aggarwal [22] introduced a feature-based stereo algorithm with zero-crossing as a matching feature. Figural continuity was incorporated into the algorithm by using zero-crossing patterns. Each edge pixel with its 8-neighboring possible edge points were classified into nine 3×3 patterns according to their vertical connections. A relaxation technique is used with initial local probability obtained from the similarity of matching patterns and the difference in intensity gradients. The probabilities are iteratively improved by using the continuity of the disparity. No result for repetitive scene patterns or scenes with occlusion is given.

Baker and Binford [31] proposed a similar algorithm. They incorporated the coarse-to-fine strategy, but they used a modified Viterbi algorithm to match the edges between each pair of the epipolar lines. The Viterbi algorithm is a recursive optimal solution to the problem of estimating the state sequence of a discrete-time finite-state Markov process. Baker and Binford assumed that no edge reversals occur in the image plane. Therefore, in their assumption the same edge sequence in the left image will occur in the right image plane. The Viterbi technique is different from the normal search methods because it partitions the original problem into two sub-problems recursively each of which can be solved optimally. The Viterbi algorithm is implemented by an array of $p(i\ ,\ j)$ where i and j represent the ith and jth edges in the left and the right image. Each entry in $p(i\ ,\ j)$ has associated with it a local score, a cumulative score, and predecessor links. In the reduced resolution, the local score is evaluated from edge point attributes, such as contrast about the edge, intensity difference about

the edge, and interval compression ratio. In the case of the full resolution, orientation of the edge and the reduced resolution correspondence probabilities are also included. The problem with this technique is that it is computationally very expensive because at each edge pixel the probability scores have to be evaluated.

Ohta and Kanade [73] extended Baker's method to include the inter-scan line search in order to exploit the edge continuity constraint (figural continuity). A sub-optimal global match is achieved by this technique although it is computationally very intensive. One major problem with the Viterbi algorithm is that the order of edges in the left and the right image must be preserved.

Recently Medioni and Nevatia [20] introduced a curve-based matching technique. In their algorithm local edges were extracted by using the 5×5 directional Nevatia-Babu edge operators [38]. After thinning and linking the edges, edge boundaries were segmented into piece-wise linear segments. For each linear-edge segment some local features were extracted, such as the length of the edge, average contrast along the edge and orientation of the edge segment. A measure of match between the left and the right edge segments were evaluated within a window representing the maximum disparity. This measure of match also incorporated a global match between all the possible matches within this window. Therefore a global consistency between possible matches was evaluated.

Ayache and Paverjon [57] introduced a similar technique where linear-edge segments lying within a window were considered for match. Local predictions for tentative matches were hypothesized between segments which intersect a common epipolar plane, and whose disparity lies within a predefined window (maximum disparity). A global verification is then performed for each hypothesis in a recursive manner. This is achieved by assigning new matches between neighboring segments which intersect a common epipolar plane whose disparity is close to the disparity computed between the previously matched segments, and which verify some loose geometrical similarities.

Hwang and Hall [58] introduced a stereo matching technique where relational tables for the left image and right image were formed for solving the correspondence problem. Images were first segmented into regions. Edge segments and vertices were

extracted and labeled. The structural relationships among these labels in each image were tabulated in a relational table. A global match between the two relational table were then formed.

In the gradient techniques, the stereo images are registered with each other by using some numerical techniques, such as the Newton-Raphson iteration method or the Hill Climbing iteration [45]. Spatial intensity gradient techniques have been used for the measurement of optical flow [44]. One such method was proposed by Lucas and Kanade [24]. Their algorithm starts with an initial estimate of the disparity, and it uses the spatial intensity gradient at each point of the image to modify the current estimate of the disparity. This process is repeated in a kind of Newton-Raphson iteration. Problems with this technique are that a number of iterations have to be perfo:med, and a good initial estimate for disparity is needed to start up the iteration. However, sub-pixel accuracy is possible with this technique. It would be very interesting if gradient techniques and edge-based matching methods could be combined.

Barnard and Thompson proposed [26] an algorithm to solve the disparities between two images caused by binocular or motion parallax. In their algorithm they extracted a large number of distinct features using Moravec interest operator, (the sums of the squares of the differences of pixels in four directions are computed over a small area), in the left and the right image. For each potential candidate point in the left image, an initial collection of possible matches from the right image that lie within a given distance are established. A probability confidence based on intensity difference around each match point is determined. The estimates are iteratively improved with a relaxation labeling algorithm that uses a continuity constraint. Using the continuity constraint, the candidate point will receive a support from its neighboring matched candidate point if they have approximately the same disparity and are within a proximate distance. Barnard and Thompson demonstrated their algorithm on a stereo image and showed that after 8-10 iterations almost all the candidate points in the left match their corresponding points in the right image. This technique is only successful if a large number of target points are considered. In order for the continuity constraint to work, the target points have to be dense and close to each other.

Recently a new technique was proposed [32] only for planer objects where no point correspondence was used. Lin and Binford [33] also introduced a new technique where junctions in a pair of stereo images were tried to be matched with a global constraint. A correspondence is said to be obtained for a junction if the junctions connected to it are matched, and the iconic properties of the curve joining the two junctions are satisfied. In [80] Lin and Binford introduced a hierarchical stereo vision system. Bodies, surfaces, curves, junctions, and edgels are used as feature primitives in the matching process. In their algorithm, bodies are matched followed by surfaces and then junctions, and then curve-segments joining them. Thus, high-level features are matched first and the resulting constraint are used to match low-level features. Hoff and Ahuja [77] have recently introduced a stereo vision algorithm where feature matching and surface interpolation was integrated.

3. The Overview of the Proposed Curve-Segment Stereo Matching

In this report, a new stereo vision technique is proposed [34] to solve the correspondence problem. Given a pair of stereo images, several Laplacian of a Gaussian filters ($\nabla^2 G$) of different filter size are applied to each image to extract features at different frequency ranges (channels). For each channel the zero-crossings for the left and the right image are extracted. A tracking algorithm is then used to split the zero-crossings into curve-segments. For each curve-segment the centroid is calculated using the coordinates of the edge points that form the curve-segment. The Generalized Hough Transform (R-table) [46] of the curve-segment is evaluated using the centroid as the reference point. Once the Generalized Hough Transform is evaluated for all the curve-segments in the left and the right image, the matching process looks for instances of the left curve-segments in the right image.

To obtain a global match, the structural information about the objects in the scene must be used in the matching process. To achieve this a multi-channel graph matching technique is proposed. At each channel a graph is formed from the curve-segments in the left image, where the centroids of the curve-segments represent the nodes of the graph, and the extracted information about the curves represents the local

properties of the nodes. A similar graph is formed from the curve-segments in the right image. In a pair of stereo images the structural information about each individual object is almost preserved, especially when the distance between the two cameras is very small. Thus a graph isomorphism (clique finding) technique is used to find the best subgraph match between the left and the right image graphs. Also a relaxation technique [56] with the help of the epipolar constraint on the centroids is used to find the corresponding nodes between the two graphs. Since each object in the scene can have different disparities a pure graph isomorphism is not always possible. However, the relaxation technique will find the best sub-graph in the left image that matches the graph in the right image.

The matching process starts at the coarsest channel, and the matching procedure is repeated for each channel. The disparity information from the coarser channels is used to converge the correspondence for the curve-segments from the finer channels. At the coarsest channel a "booting" disparity value is used to begin the process.

Once the curve-segment correspondence has been achieved, the pixel disparity is evaluated by subtracting the x-coordinates of the edge pixels in the left curve-segment from the x-coordinates of the right curve-segment (due to the camera's geometry, the disparity is only a horizontal displacement in the x-direction). This pixel disparity information is used to converge the matching process at the finer channels. In the following subsections we discuss the details of the proposed technique.

3.1. Edge Detection

The zero-crossings in the left image and the right image of a stereo pair are detected by using a two-dimensional Laplacian of a Gaussian operator given by,

$$\nabla^2 G(i,j) = \left[\frac{(i^2 + j^2)}{\sigma^2} - 2\right] \exp\left\{\frac{-(i^2 + j^2)}{2\,\sigma^2}\right\} \qquad (12)$$

where the size of the operator and its spatial frequency characteristics are determined

by the value of the constant σ . Following Grimson's representation [29], the central width of this operator W is given by the following expression

$$W = 2\sqrt{2}\,\sigma \tag{13}$$

The operator size is limited to a window size of $1.8W_\sigma$ because the magnitude of the coefficients falling outside this window is very small. Zero crossings are obtained by scanning along each processed image line and locating pairs of adjacent elements of opposite signs.

The orientation θ_{xy} is calculated from the local gradients in the X and Y direction of the convolved image. The orientation at each edge pixel is needed in the tracking algorithm and in the Generalized Hough Transform evaluation. The local gradients in the X and Y direction are calculated by the following operators:

$$\delta Y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{14}$$

$$\delta X = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \tag{15}$$

Then the orientation is given by,

$$\theta_{xy} = \tan^{-1} \frac{\delta Y}{\delta X} \tag{16}$$

Zero crossings are obtained by scanning along each processed image line and locating pairs of adjacent elements of opposite signs. This edge detector is known as the Marr-Hilderth operator [35] which is already simulated on the computer. Figure 4 represents the zero-crossings extracted from a pair of stereo images and Figure 5 represents the corresponding orientation of the zero-crossings. Figure 6 represents the original pair of stereo images.

Two other techniques such as Haralick [36] and Canny [37] edge detectors were investigated. Haralick proposed an edge detector, in which he suggested that the underlying gray tone intensity around each image pixel can be approximated by a

cubic polynomial. This cubic polynomial was decomposed into a set of discrete orthogonal polynomials. Edge pixels were detected when there was a zero-crossing of the second directional derivative taken in the direction of the gradient. Canny studied the desirable properties of an optimal edge detector, and his criteria were based upon the efficiency of detection and reliability in localization. He designed an optimal edge detector, (according to his criteria), which approximated to finding of maxima in gradient magnitude of a Gaussian-smoothed image. Torre and Poggio [39]-[40] introduced an edge detection technique based upon a regularization theory. In this method the edge detection is considered as an ill-posed problem, and the solution was obtained by finding a filter that minimize an appropriate functional.

3.2. Feature Primitive and Curve Tracking

We have chosen curve-segments as our matching feature primitive because of three desirable characteristics. First, a curve-segment can still be identified even if it is partially occluded. Also, the figural continuity constraint is automatically satisfied by using curve-segments. Finally, the problem of finding a unique match for a curve-segment is much less ambiguous than for a point target and should produce fewer mismatches.

One advantage of using curve-segments as a feature primitive over edgels is that a graph can be formed for each image to represent the local properties of the curve-segments as well as the relational (structural) properties between the curve-segments. Consequently, a graph matching technique (relaxation) may be used to obtain a global match between the left image curve-segments and that of the right image curve-segments. It is also important to note that the centroids of the curve-segments should satisfy the epipolar constraint imposed on them by the camera's geometry.

To identify each curve-segment, the Generalized Hough Transform representation of the curve is evaluated. This represents the iconic property of the curve-segment and is used to identify the instance of the same curve-segment in another image. Other image signatures such as Curvature vs Arclength or Slope Density Function [47] could have been used to represent the iconic properties of the edge pixels. The Generalized

Hough Transform can easily be extended to include other properties of the curve, such as the average gray level, gradient, curvature, color, end point locations of the curve-segment, or the type of junctions terminating the curve-segment.

A tracking algorithm [55] is used to segment the boundaries of the objects in the scene into short curve-segments. Each curve-segment is labeled by a number, the locations of all the edge points forming the curve-segment are stored, and the centroids of each curve-segment are calculated. The starting points for tracked curve-segments are the edge pixels whose gradient magnitudes are above a pre-defined high threshold. From each starting point the curve is tracked recursively in both directions by considering the magnitude and the orientation of the neighboring points. At any point on the curve each of its 8-neighborhood points is included in the tracked curve if its gradient magnitude is above a pre-defined low threshold, and its orientation does not differ by more than a pre-defined threshold, say 10°, from the previously tracked point. If the difference in the orientation between two tracked points exceeds the threshold, or if the magnitude of the new point dips below the low threshold, then the curve is broken off. Also as each curve is tracked, a record of the edge pixel locations is kept, as well as the total length of the curve-segment; very short curve-segments are discarded. A set of curve-segments is thus obtained for the left and the right image. For each curve-segment its centroid is calculated which is given by

$$\bar{c}_x = \frac{1}{N} \sum_{i=1}^{N} x_i \tag{17}$$

$$\bar{c}_y = \frac{1}{N} \sum_{i=1}^{N} y_i \tag{18}$$

where N is the length of the curve-segment and (x_i, y_i) the coordinates of the edge pixels. The above information is then used in the Generalized Hough transform algorithm to generate the distinctive R-table of each curve. Figure 7 shows the curve-segments extracted from a pair of stereo images. The curve centroids are represented by cross-marks.

More research is needed on the tracking algorithm, the simulated technique in this proposal is not always effective and sometimes it will produce broken curve-segments. The

parameters needed to obtain the same curve-segments for each image may be different. We intend to investigate better ways of tracking and segmenting the zero-crossings. For example, we could track curves and segment them at the points of high curvature.

3.3. The Generalized Hough Transform

To identify an instance of each left curve-segment in the right image we must extract and use some distinctive features about the curve. We use the Generalized Hough Transform [46] of the curve as a distinct property of the curve-segment. The Generalized Hough Transform is used because it groups all the information about the individual zero-crossings on the curve into a single table which uniquely represents that curve-segment. The Generalized Hough Transform is very robust in finding a match for a curve-segment even when some part of the curve is occluded, or when edge pixels are noisy and point-to-point matching is not possible. A table called the R-table is constructed for each curve-segment in the left and the right image. This is done by calculating the orientation θ_{xy} of each edge pixel (e_x^i, e_y^i) of the curve-segment E^i and the vector distance between the centroid of the curve-segment $C^i = (c_x^i, c_y^i)$ and the edge pixel location; this distance R^i is given by $(r_x^i, r_y^i) = (c_x^i - e_x^i, c_y^i - e_y^i)$. In the R-table these (r_x^i, r_y^i) are listed as a function of θ_{xy}. This table is used to detect instances of the same curve segment in the right image, with the additional constraint that the difference in the locations of the matching centroids is only a horizontal shift. This difference in location is a result of the parallel axis camera geometry. A curve-segment in the left image is said to be matched with a curve-segment in the right image if their R-tables are approximately the same. Other properties of the curve, such as the average gray level, average gradient, average curvature of the edgels on the curve, or color of the curve, may also be used to represent the local characteristics of the curve and be used as constraints in the matching process.

3.4. Matching Process

The tracking algorithm produces a set of curves with distinct local characteristics, stored in the R-tables of the curves. The curve stereo matching problem may now be considered as a point matching problem between a set of points from the left image and a set of points from the right image. Since the number of centroids is much less than the number of edge pixels, the matching is easier and faster. Also sophisticated matching techniques exist in which the relationships among the curve centroids may be used to obtain a global match between a set of centroids in the left and the right image.

A multi-channel graph matching technique is proposed to obtain a global match between curve-segments in the left image and those in the right image. In this technique, two graphs are formed in each channel where the graphs represent the structural relationships and the local properties of the centroids of each image. Sub-graph isomorphisms are then obtained by a relaxation technique or a graph isomorphism method (Clique finding) for each channel.

3.4.1. Multi-Channel Graph Matching

The input to the matching algorithm is the set of curve-segments extracted from the left and the right image. Each curve-segment is identified by a number, the location of its centroid, its R-table representing the properties of the curve-segment, its curve-length, and the location of each edge pixel belonging to the curve.

Using the centroids and the R-tables of the curve-segments, a relational graph can be formed for each image. The nodes of the graph represent the location of the centroids of the curve-segments, and the arcs represent the relationship between the centroids. The R-table of the curve-segment represents the local (iconic) properties of the node and the distances between the centroids are used to represent the structural characteristics of the objects in the scene.

Let graphs $L(N_l, P_l, E_l, \sigma)$ and $R(N_r, P_r, E_r, \sigma)$ represent symbolically the left and the right image respectively where N represents the number of nodes (curve-

segments), P a set of local properties of the nodes, E a set of relations between the nodes, and σ representing the channel. In our matching algorithm, the R-table of each curve-segment is the local property of the corresponding node, and the distances between the nodes are the relational properties between the nodes.

If local properties P_l^i of a node N_l^i in graph L approximately matches, by a similarity measure, the local properties P_r^j of a node N_r^j in graph R, then this pair of nodes (N_l^i,N_r^j) is said to form a node assignment, provided the epipolar constraint on the centroids is also approximately satisfied. The measure of local similarity for node assignment in our matching algorithm is given by the ratio

$$S(N_l^i,N_r^j) = \frac{A^j}{L_l^i} \tag{19}$$

where A^j represents the Hough accumulator value obtained when the R-table of the left curve-segment E_l^i is compared with the R-table of the right curve-segment E_r^j and L_l^i represents the curve-length. Two node assignments (N_l^i,N_r^j) and (N_l^m,N_r^n) are said to be compatible if their relational properties are satisfied. A compatibility measure between them is given by the ratio

$$C(N_l^i,N_r^j\ ;\ N_l^m,N_r^n) = \frac{1}{1 + \dfrac{(d_l^{im} - d_r^{jn})^2}{B}} \tag{20}$$

where d_l^{im} represents the distance between the centroids of the *ith* and the *mth* curve-segment in the left image and d_r^{jn} that of the *jth* and the *nth* curve-segments in the right image, B represents a constant say $B = 10$. The relational graphs obtained from the left and the right image can be matched by a clique finding technique [28] or by a relaxation technique using the node assignment and compatibility measures as discussed below.

1) Epipolar constraint on the centroids: Starting at the coarsest channel for each curve-segment E_l^i in the left image a matching curve-segment E_r^j is sought in the right image in order to form a node assignment. Due to the cameras' geometry, the location $C_l^i = (c_x^i,c_y^i)$ of the centroid of the left curve-segment will be displaced by a disparity value in the right image. Ideally, the disparity value is simply a horizontal shift, i.e.,

the centroids should be on the same epipolar line. Although, due to the geometrical distortion, imperfections in the tracking of the curves, and partial occlusion, the centroids of the corresponding curves may not be exactly on the same epipolar line. However, for most of the corresponding curve-segments their centroids will fall within a few scan lines of each other. The horizontal displacement between the centroid of the left curve-segment and that of the corresponding right curve-segment is called the centroid disparity.

2) Node Assignment: To find the node assignments for a given curve-segment E_l^i in the left image all the nodes (centroids) that are within a search window ($2W_\sigma \times W_{\sigma=1.5}$) around the point $(c_x^i + \bar{d}^i, c_y^i)$ in the right image are considered as candidates. Here, (c_x^i, c_y^i) is the location of the node (centroid of the left curve-segment) and $\bar{d}^i$ is the average disparity around the curve-segment which is obtained from the previous channel disparity buffer. An initial rough estimate of disparity is assumed for the coarsest channel.

The R-table of each candidate is compared with that of the left curve-segment and an initial score between the two curve-segments is calculated. To compare the R-table of the left curve-segment with that of the right curve-segment, each θ_{xy}^i entry of the R-table of the left curve-segment is compared with all the θ_{xy}^j entries of the R-table of the right curve-segment. If the orientations are approximately the same $|\theta_{xy}^i - \theta_{xy}^j| \leq 5^o$ then the location $((c_x^i + r_x^i - r_x^j) , (c_y^i + r_y^i - r_y^j))$ in the two-dimensional Hough accumulator is incremented by one. Thus, if there is a similarity between the R-tables there will be a Hough peak. Let the value A^j represent the Hough peak (number of matched points) at the location (c_x^i , c_y^i) in the Hough accumulator obtained from the curve-segment j. The local node assignment score is given by the expression (19). If there is no candidate with a node assignment score satisfying

$$S(N_l^i, N_r^j) = \frac{A^j}{L_l^i} \geq 0.7 \tag{21}$$

then this left curve-segment has to go through a modified node assignment process which is discussed below.

3) Node assignment for occluded and broken-distorted curves: Due to occlusion,

imperfect tracking, and distortions in image formation, some of the curve-segments extracted from the left and the right image do not have the same curve-length. This can cause the centroids from the right image to fall outside of the search windows described above, and can result in a failure to match the curve from the left image. This problem is especially serious for the finest channel when the window size is very small.

Any curve-segment in the left image may find a match with one or several curve-segments in the right image, (i.e. broken curve-segments), and vice versa. For each of these unmatched curves we perform a standard Hough Transform against all the curve-segments in the right image. Any curve-segment j in the right image that gives a peak value A^j within a search window of size ($2W_\sigma \times W_{\sigma=1.5}$) around the displaced centroid of the left curve-segment $(c_x^i + \bar{d}^i, c_y^i)$ in the Hough accumulator buffer, is said to be a matching curve and the node assignment is given by

$$S(N_l^i, N_r^j) = \frac{A^j}{Minimum\ (L_l^i, L_r^j)} \tag{22}$$

For each matching curve a new curve-segment is created with the location of the Hough peak as its new centroid, and a new R-table based upon this centroid is formed. If there are two matching curves with their centroids a few pixels from each other then they are combined into one single curve-segment and one of the centroids is assigned to it. Table 1 shows the centroid locations of all the curve-segments extracted and created from the left and the right stereo images of Figure 9. Curve-segments #40 to #72 were generated during the node assignment process.

There is a possibility that there is more than one node assignment for a given left curve-segment within the searching window of size ($2W_\sigma \times W_{\sigma=1.5}$) in the right image. To resolve the ambiguity in false matches the structural compatibility between node assignments has to be employed by a global matching technique.

3.4.2. Clique Finding

The correspondence between two sets of nodes N_l and N_r obtained from the left and the right image respectively can be considered as a point pattern matching or a graph matching problem. To match two relational graphs we have to find the matching nodes, two graphs are said to be isomorphic if there exits a one-to-one node assignments which are also mutually compatible. Due to occlusion and noise a complete match (isomorphism) between the two relational graph can not be found. But a subgraph of the left image graph can find a match with the subgraph of right image graph this isomorphism is known as "Double" subgraph isomorphism". A well known technique to find all the matching subgraphs is the clique finding problem. A clique is a subgraph that is totally connected (nodes are said to be connected if they satisfy the compatibility measure). To match the two graphs first an association graph is formed. An association graph is a graph with nodes consisting of an assignment of a pair of nodes from the left and the right image graphs. The maximal clique of the association graph will give the largest subgraph match between the left and the right relational graph. A procedure to find the maximum clique is given in [47] and [50] which was simulated to find the best corresponding points [90]. It was found that when there are several objects with a large depth difference between them, the compatibility measure between the objects will not be satisfied, thus resulting in a maximal clique of a very small size. In the next section, a relaxation technique is introduced which will produce more matches than the maximal clique finding approach.

3.4.3. Relaxation

Ranade and Rosenfeld [34] developed a fuzzy relaxation technique for point matching where matching scores were updated by the expression

$$S^{(r+1)}(N_l^i, N_r^j) = \frac{1}{n} \sum_{N_l^m \in K}^{N_l} \sum_{m=1}^{N_l} \left[\max_{n=1}^{N_r} C(N_l^i, N_r^j; N_l^m, N_r^n) \; S^{(r)}(N_l^m, N_r^n) \right] \quad (23)$$

where $C(N_l^i, N_r^j; N_l^i, N_r^n) = 1$ if $j = n$ and 0 otherwise. In this relaxation technique the initial score $S^o(N_l^i, N_r^j)$ is updated by the maximum support from neighboring pair of

nodes. For each node N_l its node assignment score is updated; only the nodes that form a node assignment and are within the neighborhood distance of K pixels from it can contribute to its node assignment score. The pair of nodes that have the same disparity will contribute significantly and the nodes that have different disparities will contribute very little. As the iteration is performed the node assignment score is decreased; however, the score decreases faster for the less likely matche han for the most likely ones. Table 2 represents the initial and the final node assignment scores for each of the left curve-segments shown in the Table 1. For each curve-segment, the candidate with the maximum score after several iterations is chosen as the most likely corresponding curve-segment. Table 3 represents the matched curve-segments for the candidate curve-segments shown in Table 1 for the first channel $\sigma = 6.0$, the centroid disparity is also given.

3.4.4. Size of the Searching Window and the false target

The decision on the size of the searching window (to avoid false targets) that is allowed was investigated by Marr-Poggio in [25]. It was shown that the probability distribution of the interval between adjacent zero-crossing of the same sign depended on the image characteristics and the filter characteristics.

The stereo images were convolved with bandpass Gaussian filter of central width W_σ . For a given zero-crossing in the left image the probability of another zero-crossing of the same sign in the right image was less than 5% if the disparity range over which a match is sought was restricted to $\pm\frac{1}{2}W_\sigma$. However, such a disparity range is very restrictive especially when there are several objects with large disparity differences (it is very costly to use a larger filter size). Thus, Marr and Poggio investigated a larger searching area of $\pm W_\sigma$. However, this situation produced a probability of false targets of about 50%. Therefore, 50% of all the possible matches will be ambiguous that is they will have multiple matches.

In our algorithm for a given curve-segment in the left image the probability of another curve-segment of the same sign and the same R-table in the right image within

a search window of $\pm W_\sigma$ is much less than 50%. This was pointed out by Mayhew and Frisby (figural continuity constraint). So our search window of size $\pm(W_\sigma \times 7.5)$ or larger is appropriate. However, in situations where there are repetitive patterns and occlusion of multiple objects the probability of a false target will increase. To resolve the ambiguities in multiple matches, the relaxation technique uses the structural information about the scene to find the match with the highest score.

3.4.5. Coarse-To-Fine Control Strategy

To bring the curve-segments obtained from the finer channel into correspondence, the pixel disparities from the coarser channels are used. This is done for each curve E_l^i by finding the average disparity in a region (about $2W_\sigma$ pixels width with the σ of the previous channel) around the curve-segment in the disparity buffer from the previous (coarser) channel. The average disparity $\bar{d}^i$ is used to find the center of the search window which is given by $(c_x^i + \bar{d}^i, c_y^i)$ where (c_x^i, c_y^i) is the location of the left curve-segment centroid. We assume that disparity changes occur somewhat smoothly over the object. At depth discontinuities the average disparity does not represent the true disparity of the curve-segments since it is averaged over several curve-segments with different disparities. A better coarse-to fine strategy is needed, such that it can find the true average disparity at the depth discontinuities. One approach would be to perform a surface interpolation using the disparity of the edges and then perform a segmentation of the disparity data to locate depth discontinuities.

3.4.6. Actual Pixel Disparity

To obtain the actual pixel disparities between the pixels of two matched curve segments, we simply subtract the x-coordinates of corresponding pixels, i.e., pixels that have the same y-coordinates. There are two cases where it is difficult to form a one-to-one pixel correspondence. The first is when part of the curve is occluded. The second is when several pixels on a curve have the same y-coordinates as in horizontal lines. In these cases, we set the pixel disparities equal to the curve's centroid disparity. This is an advantage of the proposed technique compared to the local stereo matching

techniques [30] - [31] where horizontal lines are ignored. All the computed pixel disparities are stored in a disparity buffer for each channel.

4. EXPERIMENTAL RESULTS

Results are presented for running the algorithm on a set of real images. The stereo images were taken using a single camera which was translated horizontally to obtain the left and the right images. Each stereo pair was of resolution 256 × 256 and 8 bits gray level. The zero-crossings were extracted after convolving the stereo images with $\nabla^2_\sigma G$ for three σ values ranging between 1.5 to 6.0. This range for σ does not represent the actual size of the channels in the Human Visual System, but this range is adequate for experimental demonstration of the algorithm.

Figure 9 shows stereo images obtained at three different camera positions when it was translated horizontally. Figure 10 shows the zero-crossings for the left and the middle stereo images. The curve-segments for the left, the middle and the right stereo images are shown in Figure 11. The disparity between the left and the middle stereo images as well as the disparity between the left and the right stereo images are shown in Figure 12. In these stereo images there are three objects with different disparities and there are partially or totally occluded curve-segments. Centroid disparity is assigned to the horizontal curve-segments as well as to the regions of the partially occluded curve-segments. In Figure 13 we have identified and labeled a few of the curve-segments at $\sigma = 6.0$ for the left and the right stereo images. For example the curve-segment #9 in the left image will match with the curve-segment #49 in the right image with a centroid disparity of $d = 51$ as shown in Table 3. From Table 2, it is also found that the curve-segment #9 will form a match with the generated curve-segments #50, #51. However after the relaxation process the node assignment scores for these matches are smaller than that of the curve-segment #49, therefore these correspondences are assumed to be mismatches. Curve-segment #9 should have actually formed a node assignment with the curve-segment #9 in the right image, but because the match criteria given by expression (21) was not satisfied three new curve-segments were generated as possible candidates. These three new curve-segments #49,

#50, and #51 were formed when curve-segment #9 in the left image was matched against curve-segments #9, #25 and #34 in the right image respectively. Let us now consider the curve-segment #6 in the left image, this curve-segment is almost completely occluded in the right image, following the discussion in section 3.4.1 the matching process will try to find a match for this curve-segment by creating a new curve-segment #44 from the occluded curve-segment #8 in the right image. This match will result in a centroid disparity of $d = 54$ approximately the same as that of the curve-segment #9 since they belong to the same object. The horizontal curve-segment #2 in the left image will find a perfect match with the horizontal curve-segment #2 in the right image with a centroid disparity of $d = 68$. Similarly the curve-segment #17 will find a perfect match with the curve-segment #16 with a centroid disparity of $d = 23$. It also generates a match with the curve-segment #10, but after relaxation process this mismatch is identified and discarded. Figure 13 also shows the curve-segments #18, 20, 21, 22, 26, 28, 35, 14, 38, 10 and their corresponding curve-segments #20, 22, 26, 23, 30, 32, 36, 19, 37, 12 respectively. In Table 3, it is seen that curve-segments #2, 7, 11, 18, 20, 21, 22, 26, 27, 28, 29, 34, 35, 37, and 39 belong to the Diet Coke Can located in the middle of the stereo images with an expected average disparity of $d = 70$. Curve-segments #1, 4, 6, 9, 12, 13,14, 24, 25, and 38 belong to the Sunkist Can located in the far left of the stereo images with an expected average disparity of $d = 52$. Curve-segments #4, 12, and 25 belonging this object are not matched correctly, because these curve-segments do not exist in the right image. Curve-segments #5, 8, 10, 17, 19 belong to the Coca-Cola Can located in the far right of the stereo images with an expected average disparity of $d = 23$. A pair of stereo images with a repetitive pattern is shown in Figure 14. The zero-crossing, curve-segments and their disparities are shown in Figure 15 16 and 17 respectively. Since a global match was achieved by the relaxation technique stereo images with repetitive patterns can be fused. Figure 7 shows the extracted curve-segments for the stereo images of Figure 6 for different channels, in this pair of stereo images there is a vertical disparity of three pixels due to misalignment of the cameras. Figure 8 represents the disparity image.

5. CONCLUSION AND DISCUSSION

A curve-segment based stereo vision algorithm has been presented. Curve-segments are used as the feature primitives in our matching process. Uniqueness of matching is enforced by the inherent figural continuity property of the curve-segment and the disparity similarity between the curve-segments. The R-table of the Hough Transform of each curve-segment in the first image is used to form node assignments with all the possible candidates in the right image. The relaxation technique uses the global consistency between the curve-segments (similarity in disparity) to disambiguate the false matches.

The algorithm discussed in this paper is similar to the Marr-Poggio-Grimson stereo vision technique. The difference is that curve-segments are used as the matching primitive rather than zero-crossing points. Also, a relaxation technique is used to resolve ambiguous matches rather than the pulling strategy proposed in the MPG stereo algorithm.

The significance of the proposed algorithm compared with the current stereo vision algorithms is that, the correspondence problem between edge pixels has been reduced to that of finding correspondences between two sets of nodes (centroids). Since the number of the target points, (the curve-segment centroids), is much less than the number of the edge pixels, high level matching techniques can be used to solve the correspondence problem. Geometrical properties of the curve-segments and the relation among the centroids are used as constraints to guide both the local matching and the global matching process in resolving the ambiguities in multiple matches.

The proposed algorithm can be considered as an improvement of the Marr-Poggio-Grimson stereo algorithm; we have extended their edgel matching primitive to curve-segments which drastically reduces the number of mismatches. Also the ambiguity in resolving multiple matches is solved by using the relational information between the neighboring curve-segments. The disparity range allowed is also much larger than that in the MPG algorithm this is discussed in section 3.4.4. Partially occluded curve-segments can also be matched and the region on the curve-segment that are occluded assume the centroid disparity of the curve-segment.

The proposed algorithm is believed to improve the state of the current stereo vision techniques. For example, local matching and global matching are incorporated into the stereo system. It is a feature-based technique, and the figural continuity constraint is an inherent property of the algorithm. Apparent curve-segments, due to illumination variation, will not find correspondence because they fail to satisfy the relational constraints. Using high level matching techniques, occluding and vanishing edges can be identified.

6. REFERENCES

[1] Kender, J., "Shape from texture", Tech. Rep. CMU-C5-81-102, Dept. Computer Science, Carnegie-Mellon Univ. Pittsburgh, PA, 1980.

[2] Horn, B. K. P., "Robot Vision", The MIT Press, 1986.

[3] Shafer A. S., "Shadows and Silhouettes in Computer Vision", Kluwer Academic Publishers 1985.

[4] Steven, K. A., "The Visual Interpretation of Surface Contours", in Computer Vision edited by J. M. Brady.

[5] Shirai, Y. and Suwa M., "Recognition of Polyhedrons with a Range Finder", in Proc. 2nd Int. Joint Conf. Artificial Intell., London, Sept. 1971, pp. 80-87.

[6] Jarvis, R. A., "A Perspective on Range Finding Techniques for Computer Vision", IEEE Trans. Pattern Anal. Machine Intell., Vol.PAMI-5 No.2, pp. 122-139, March 1983.

[7] Thompson, A. M., "Camera Geometry for Robot Vision", Robotics Age, pp. 20-27, March/April 1981.

[8] Yakimovesky, Y. and Cunningham R., "A System for Extracting Three-Dimensional Measurements from a Stereo Pair of TV Cameras", Computer Graphics and Image Processing 7, pp. 195-210, 1978.

[9] Gibson, J. J, "The Perception of the Visual World", Boston:Houghton Mifflin, 1950.

[10] Julesz, B. "Foundations of Cyclopean Perception", Chicago:University of Chicago Press, 1971.

[11] Marr, D., "Vision", W. H. Freeman and Company, NY 1982.

[12] Richards, W., "Stereopsis with and without Monocular cues", Vision Res. 17, pp. 967-969, 1977.

[13] Hannah, M. F., "Computer Matching of Areas in Stereo Images", Stanford Artificial Intell. Lab., AIM-239, Ph.D. thesis, July 1974.

[14] Gennery, D. B., "Modeling the Environment of an Exploring Vehicle by Means of Stereo Vision", Stanford Artificial Intell. Lab., AIM-339, Ph.D. thesis, June 1980.

[15] Moravec, H. P., "Obstacle Avoidance and Navigation in the Real World by a Seeing Robot Rover", Stanford Artificial Intell. Lab., AIM-340, Ph.D. thesis, Sept. 1980.

[16] Arnold, R. D., "Local Content in Matching Edges for Stereo Vision", Proc. Image Understanding Workshop, pp. 65-72, 1978.

[17] Grimson, W. E. L., "Computing Shape using a Theory of Human Stereo Vision", Dept. of Mathematics, MIT, Ph.D. thesis, June 1980.

[18] Baker, H. H., "Depth from Edge and Intensity Based Stereo", Standford Univ., Stanford, CA., Tech. Rep., STAN-CS-82-930, Sept. 1982.

[19] Marr, D. and Poggio, T., "A Theory of Human Stereo Vision", MIT Artificial Intell. Memo No.451, Nov. 1977.

[20] Medioni, G. and Nevatia, R., "Segment Based Stereo Matching", CVGIP 31, pp. 2-18, 1985.

[21] Barnard, S. and Fishler, M., "Computational Stereo", ACM Computing Surveys 14, No.4, pp. 553-572, 1982.

[22] Kim, Y. C. and Aggarwal, J. K., "Finding Range from Stereo Images", Proc. of Computer Vision and Pattern Recognition, pp. 289-294, June 19-23, 1985.

[23] Mayhew, J. E. W. and Frisby J. P., "Psychophysical and Computational Studies towards a Theory of Human Stereopsis", Artificial Intell. 17, pp. 349-385, 1981.

[24] Lucas, B. D. and Kanade T. "An Iterative Image Registration Technique with an Application to Stereo Vision", pp. 674-679, IJCAI-81.

[25] Marr, D. and Poggio, T., "A Theory of Human Stereo Vision", MIT Artificial Intell. Memo No.451, Nov. 1977.

[26] Barnard, T. S. and Thompson, W. B., "Disparity Analysis of Images", IEEE Trans. Pattern Anal. Machine Intell., Vol.PAMI-2, No.4, pp. 333-340, July 1980.

[27] Levine, M. D., O'Handley, D. A., and Yagi, M. G., "Computer Determination of Depth Maps", Computer Graphics and Image Processing, 2, pp. 131-150, 1973.

[28] Henderson, R. L., Miller, J. W., and Grosch, C. B., "Automatic Stereo Reconstruction of Man-Made Targets", SPIE Vol.186, Digital Processing of Aeriel Images, pp. 240-248, 1979.

[29] Grimson, W. E. L., "From Images to Surfaces: A Computational Study of the Human Early Visual System", MIT Press 1981.

[30] Grimson, W. E. L., "Computational Experiments with a Feature Based Stereo Algorithm", IEEE Trans. Pattern and Machine Intell., Vol.PAMI-7, No.1, Jan. 1985.

[31] Baker, H. H. and Binford, T. O., "Depth from Edge and Intensity Based Stereo", Proc. Seventh Int. Joint Conf. Artificial Intell., pp. 631-636, Aug. 1981.

[32] Aloimonos, J. and Basu, A., "Shape and 3-D Motion from Contour without point to point Correspondence:General Principles", IEEE Int. Conf. on Robotics and Automation, pp. 518-527, 1986.

[33] Lim, M.S and Binford, T.C., "Stereo Correspondence: Feature and Constraints", Proc. Image Understanding Workshop, pp. 373-379, 1986.

[34] Nasrabadi, N. M., Liu, Y., and Chiang, J., "Stereo Vision Correspondence using a Multi-channel Graph Matching Technique", IEEE Int. Conf. on Robotics and Automation, April 25-29, 1988.

[35] Marr, D. and Hildreth, E., "Theory of Edge Detection", Proc. R. Soc. Lond. B 207, pp. 187-217, 1980.

[36] Haralick, R.M., "Digital Step Edges from Zero Crossing of Second Directional Derivatives", IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-6, pp. 58-68, Jan. 1984.

[37] Canny, J., "A Computational Approach to Edge Detection", IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-8, pp. 679-698, Nov. 1986

[38] Nevatia, R., and Babu, K., "Linear-feature Extraction and Description", Computer Graphics Image Processing, Vol. 13, pp. 257-269, 1980.

[39] Torre V., and Poggio, T. A., "On Edge Detection", IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. PAMI-8, No.2, pp. 147-163 March 1986.

[40] Yuille, A. L., and Poggio, T., "Scaling theorems for zero-crossings", MIT, Cambridge, A. I. Memo No. 722, 1983.

[41] Poggio, T., and Torre, V., "Ill-posed Problems and Regularization Analysis in Early Vision", MIT, Cambridge, A. I. Memo No. 773, April 1984.

[42] Poggio, T., "From Computational Structure to Algorithms and Parallel Hardware", Computer Vision, Graphics, and Image Processing", 31, pp. 139-155, 1985.

[43] Bertero, M., Poggio, T. and Torre, V., "Ill-posed Problems in Early Vision", MIT, Cambridge, A. I. Memo No. 924, May 1987.

[44] Horn, B. K. P. and Schunck, B. G., "Determining Optical Flow", Artificial Intell. Lab, MIT, Vol.17, pp. 185-203, 1981.

[45] Cafforio, C. and Rocca, F., "Methods for Measuring Small Displacements of Television Images", IEEE Trans. on Information Theory, Vol.IT-22, No.5, Sept. 1976.

[46] Ballard, D. H., "Generalizing the Hough Transform to Detect Arbitrary Shapes", Pattern Recognition Vol.13, No.2, pp. 111-122, 1981.

[47] Ballard, H. D. and Brown, C.M., "Computer Vision", Prentice-Hall, Inc., 1982.

[48] Rosenfeld, A., Hummel R.A. and Zucker S.W., "Scene Labeling by Relaxation operations", IEEE Trans. SMC 6, pp.420, 1976.

[49] Shapiro, L. G., and Haralick, R.M., "Structural Descriptions and Incxact Matching", IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-3, NO. 5, pp. 504-518, Sept. 1981.

[50] Nevatia, R., "Machine Perception", Prentice-Hall, Inc., 1982.

[51] Peleg, S., "A New Probabilistic Relaxation Scheme", IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-2, pp. 362-369, July 1980.

[52] Faugeras, O. D., and Berthod, M., "Improving Consistency and Reducing Ambiguity in Stochastic Labeling: An Optimization Approach", IEEE Trans. Pattern Anal. Machine Intell., Vol. PAMI-3, pp. 412-424, July 1981.

[53] Kitchen, L., and Rosenfeld, A., "Discrete Relaxation for Matching Relational Structure", IEEE Trans. Systems Man Cybernet. 9, pp. 869-874, 1979.

[54] Pavlidis, T., "Structural Pattern Recognition", Springer-Verlag, 1977.

[55] Rosenfeld, R., and Kak, A., "Digital Image Processing" Vol. 1 and Vol. 2 Academic Press, 1982.

[56] Ranade, S., and Rosenfeld, A., "Point Pattern Matching by Relaxation", Pattern Recognition, Vol.12, pp. 267-275, 1980.

[57] Ayache, N., and Paverjon, B., "A Fast Stereovision Matcher Based on Prediction and Recursive Verification Of Hypotheses", Proceedings of the third Workshop on Computer Vision Representation and Control, pp. 27-37, Oct. 13-16, 1985.

[58] Hwang, J. J., and Hall, E. L., "Matching of featured Objects Using Relational Tables from Stereo Images", Computer Graphics and Image Processing 20, pp. 22-42, 1982.

[59] Kak, A. C., "Depth Perception of Robots", in Handbook of Industrial Robotics, S.Nof, ed., John Wiley & Sons, New York, 1985, pp. 272-319.

[60] Nasrabadi, N. M., "Depth Measurement Using Stereo Vision", North American Philips Technical Report, Document No.015678, June 19, 1986.

[61] Prazdny, K., "Detection of Binocular Disparities", Biological Cybernetics, 52, pp. 73-79, 1985.

[62] Barnard, T. B., "A Stochastic Approach to Stereo Vision", Proc. of The Fifth National Conf. on Artificial Intelligence, pp. 676-680, 1986.

[63] Barnard, T. B., "Stereo Matching by Hierarchical, Micro-canonical Annealing", Proc. of Image Understanding Workshop, pp. 792-797, Feb. 1987.

[64] Hopfield, J. J., and Tank, D. W., "Neural Computation of Decisions in Optimization Problems", Biological Cybernetics, 52, pp. 141-152, 1985.

[65] Rumelhart, D. E., McClelland, J. L., "Parallel Distributed Processing", The MIT Press, Vol. 1 and Vol. 2, 1986.

[66] Wolf, P., "Elements of Photogrammetry", McGraw-Hill Publishers, 1983.

[67] Curry, S., Anderson, J. M., Baumrind, S., and Wand, B., "Stereo Camera and Stereo X-ray Devices: Comparison of Biostereometric Measurments", Vol. 51, No.10, pp. 1597-1603, Oct. 1985.

[68] Hubel, D. H., and Wiesel, T. N., "Receptive Fields, Binocular Interaction and Functional Architecture in the Cat's Visual Cortex", J. Physiol.(London), Vol. 160, pp. 106-154, 1962.

[69] Hubel, D. H., and Wiesel, T. N., "Brain Mechanisms of Vision", Scientific American, Vol. 241, No. 3, pp. 150-164, Sept. 1979.

[70] Levine, M. D., "Vision in Man and Machine", McGraw-Hill Publishers, 1985.

[71] Kuffler, S. W., Nicholls, J. G., and Martin, A. R., "From Neuron to Brain", Sinauer Associates Inc. Publishers, 1986.

[72] Kandel, E. R., and Schwartz, J. H., "Principles of Neural Science", Second Edition, Elsevier, 1985.

[73] Ohta, Y., and Kanade, T., "Stereo by Intra-and Inter-Scanline Search Using Dynamic Programming", IEEE Trans., PAMI-7, No. 2, pp. 139-154, March 1985.

[74] Metropolis, N., Rosenbluth, A. W., Rosenbluth, N., Teller, A. H., and Teller, E., "Equations of State Calculations by Fast Computing Machines", J. Chem. Phys., Vol. 21, No.6, pp. 1087-1092, June 1953.

[75] Geman, S., and Geman, D., "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images", IEEE Trans. on Pattern Anal. Machine Intell., Vol. PAMI-6, No. 6, Nov. 1984.

[76] Fisher, M. A., and Firschein, O., "Intelligence the Eye, Brain, and Computer", Addison Wesley Publishing Company, 1987.

[77] Hoff, W., and Ahuja, N., "Extracting Surfaces from Stereo Images: An Integrated Approach", IEEE The First International Conference on Computer Vision, pp.

284-294, June 8-11, 1987.

[78] Drumheller, M., and Poggio, T., "On Parallel Stereo", IEEE International Conference on Robotics and Automation, pp. 1439-1448, April 25-29, 1988.

[79] Williams, L. R., and Anandan P., "A Coarse-To-Fine Control Strategy for Stereo and Motion on a Mesh-Connected Computer", IEEE The First International Conference on Computer Vision, pp. 219-226, June 8-11, 1987.

[80] Lim, M.S and Binford, T.C., "A Hierarchical Stereo Vision System", Proc. Image Understanding Workshop, pp. 373-379, 1987.

[81] Pietikainen, M., and Harwood, D., "Depth From Three Camera Stereo" Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 2-8, June 22-26, 1986.

[82] Ito, M., and Ishi, A., "Range and Shape Measurement Using Three-View Stereo Images", Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 9-14, June 22-26, 1986.

[83] Faugeras, O. D., and Toscani, G., "The Calibration Problem for Stereo", Proceedings of IEEE Conf. on Computer Vision and Pattern Recognition, pp. 15-20, June 22-26, 1986.

[84] Bulthoff, H. H., and Mallot, H. A., "Interaction of Different Modules in Depth Perception", Proceedings of First Int. Conf. on Computer Vision, pp. 284-294, June 8-11, 1987.

[85] Mohan, R., Medioni, G., and Nevatia, R., "Stereo Error Detection, Correction, and Evaluation", Proc. of First Int. Conf. on Computer Vision, pp. 315-324, June 8-11, 1987.

[86] Blostein, S. D., and Huang, T. S., "Quantization Errors in Stereo Triangulation", Proceedings of First Int. Conf. on Computer Vision, pp. 325-334, June 8-11, 1987.

[87] Ayache N., and Lustman, F., "Fast and Reliable Passive Trianocular Stereovision", Proceedings of First Int. Conf. on Computer Vision, pp. 422-427, June 8-11, 1987.

[88] Srinivasan, R., Ramakrishnan, K. R., and Sastry, P. S., "A Contour-based Stereo Algorithm", Proceedings of First Int. Conf. on Computer Vision, pp. 677-676, June 8-11, 1987.

[89] Nasrabadi, N. M., Liu, Y., "Application of Multi-Channel Hough Transform to Stereo Vision", The Advances in Intelligent Robotics System and Computer Vision, SPIE 484, November 1-6, 1987.

[90] Nasrabadi, N. M., Chiang, J. L., "A Stereo Vision Technique Using Curve-Segments and Relaxation Matching," in Proc. of IEEE 9th Int. Conf. on Pattern Recognition, pp. 149-151, 1988.

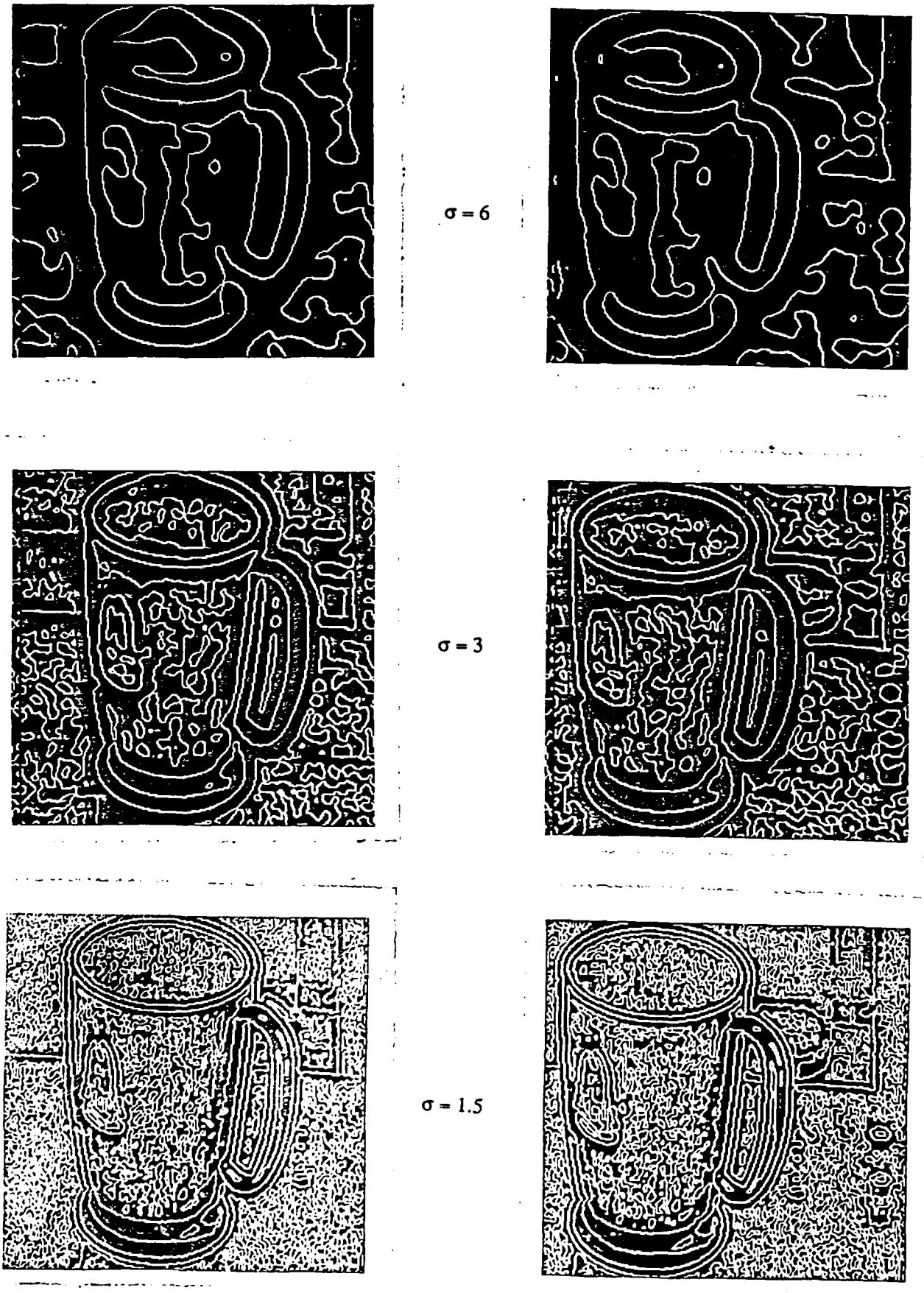


Fig. 4 Zero-crossing at σ = 6, 3, 1.5.

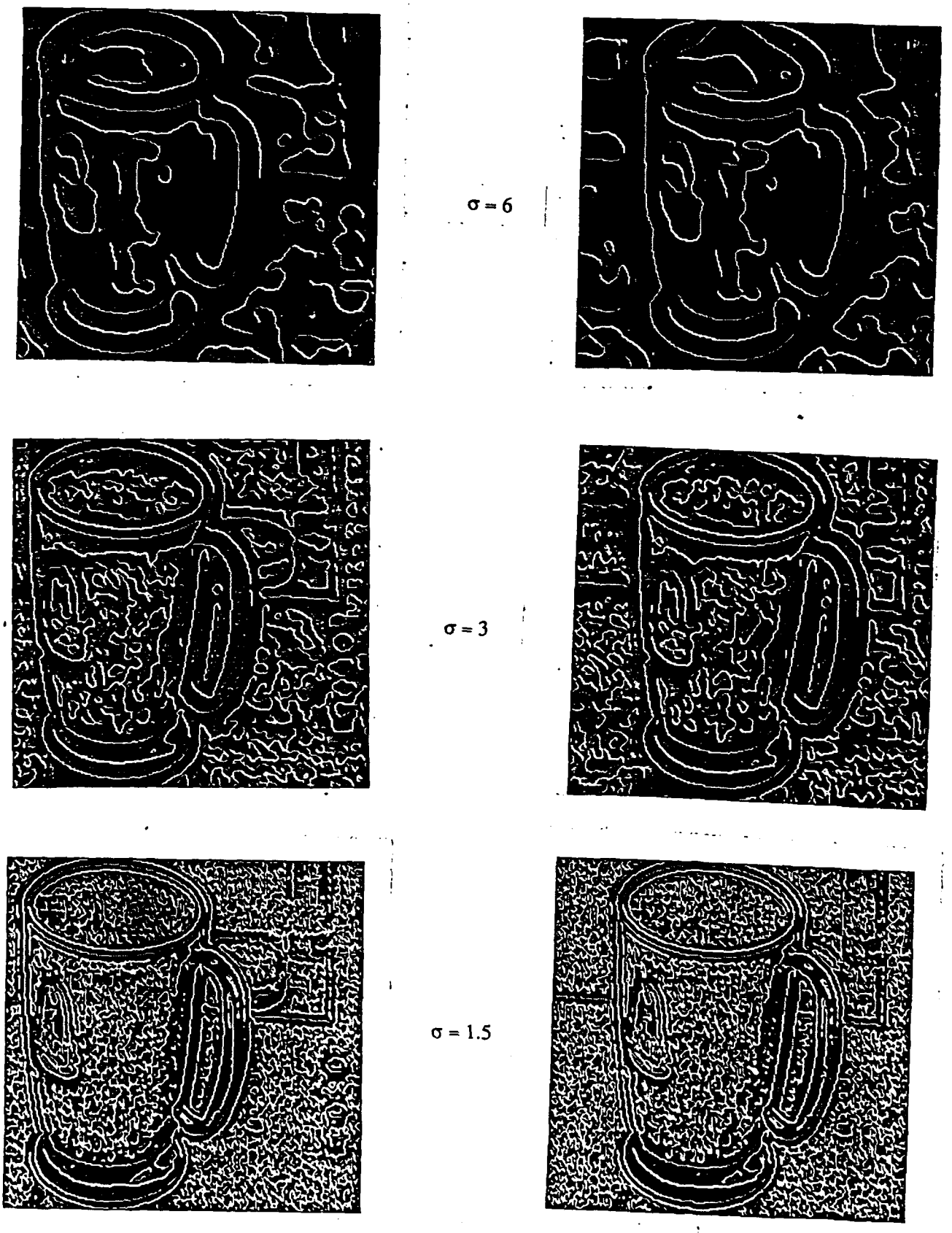


Fig. 5. Orientation of the Zero-crossings at σ = 6, 3, 1.5.

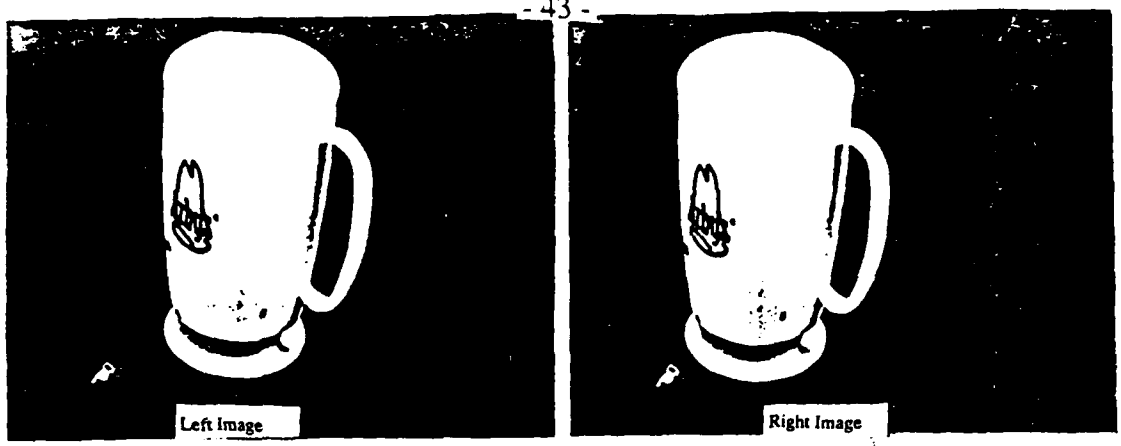
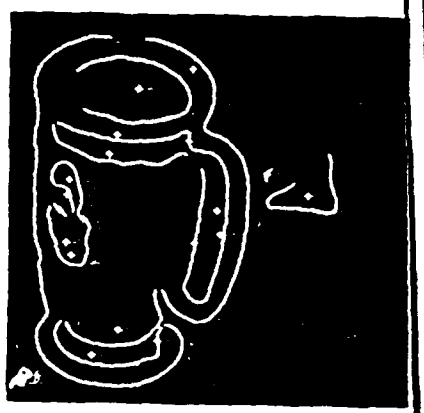


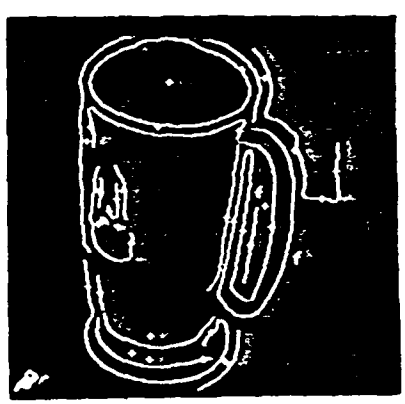
Fig. 6 A pair of stereo images of resolution 256×256 by 8 bits/pixels.

8A) Disparity at $\sigma = 6$.

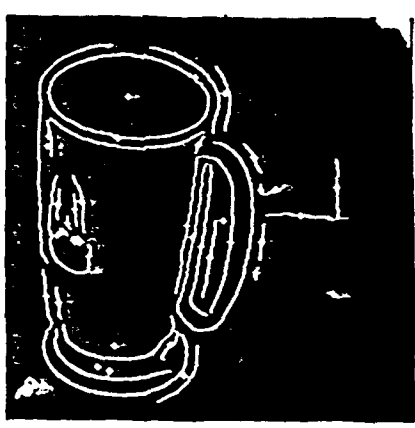
7A) Curve-segments at $\sigma = 6$.



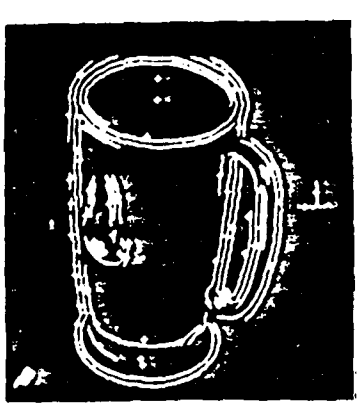
8B) Disparity at $\sigma = 3$.



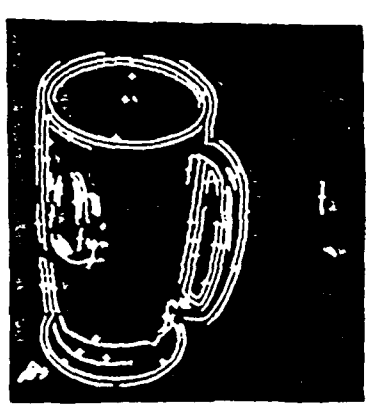
7B) Curve-segments at $\sigma = 3$.



8C) Disparity at $\sigma = 1.5$.



7C) Curve-segments at $\sigma = 1.5$.



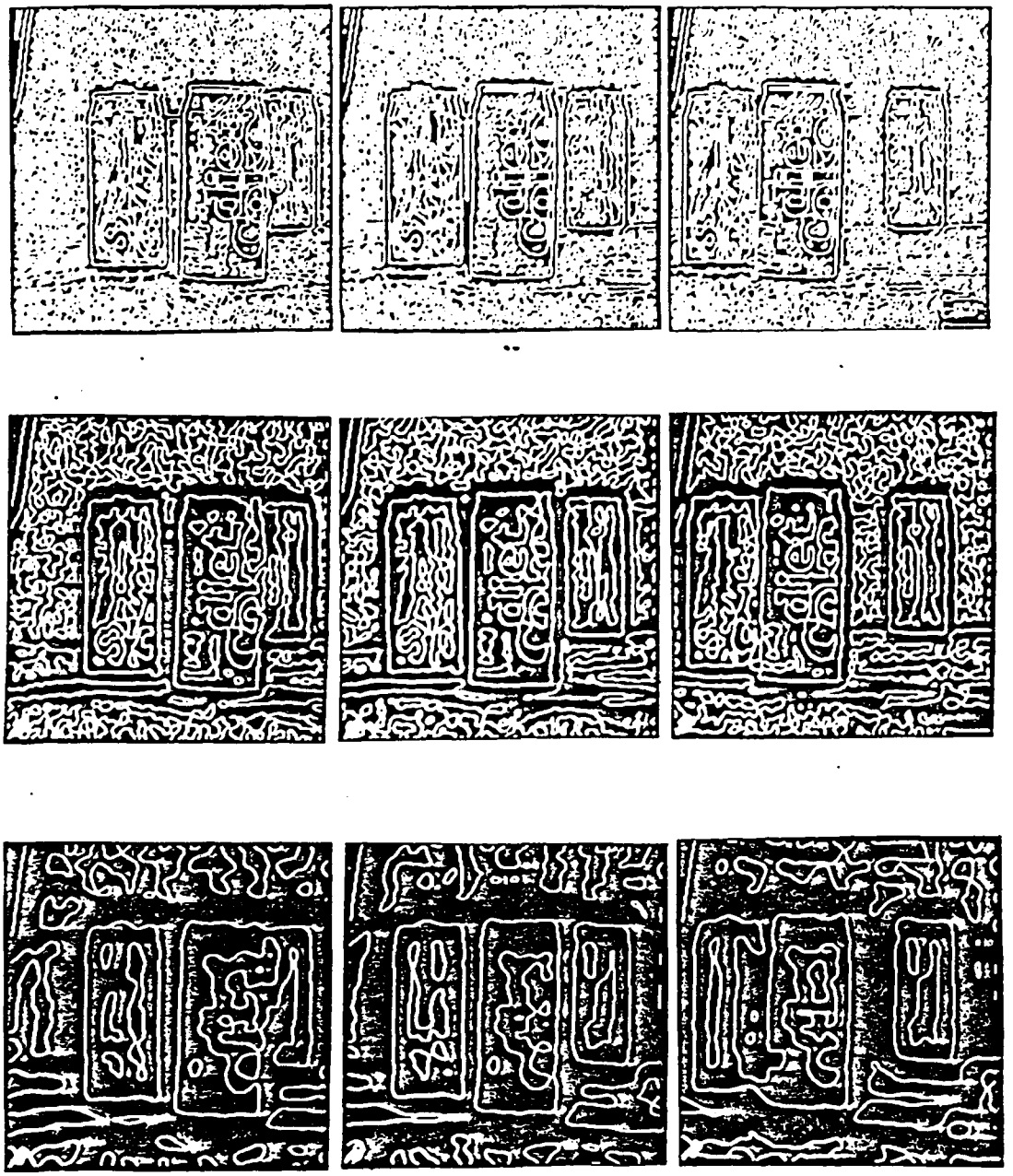


Fig. 10 Zero-crossings at $\sigma = 6, 3, 1.5$.

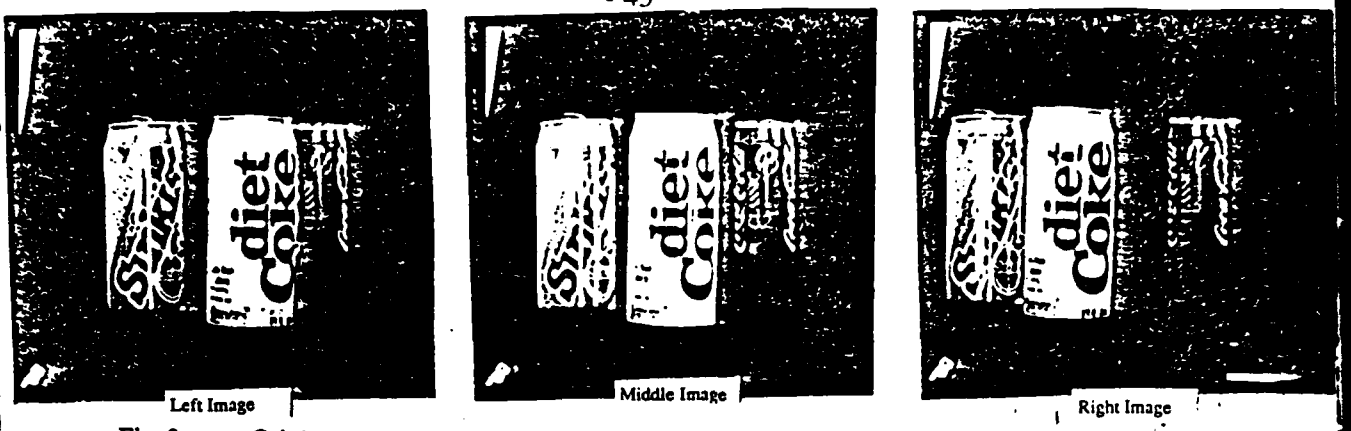
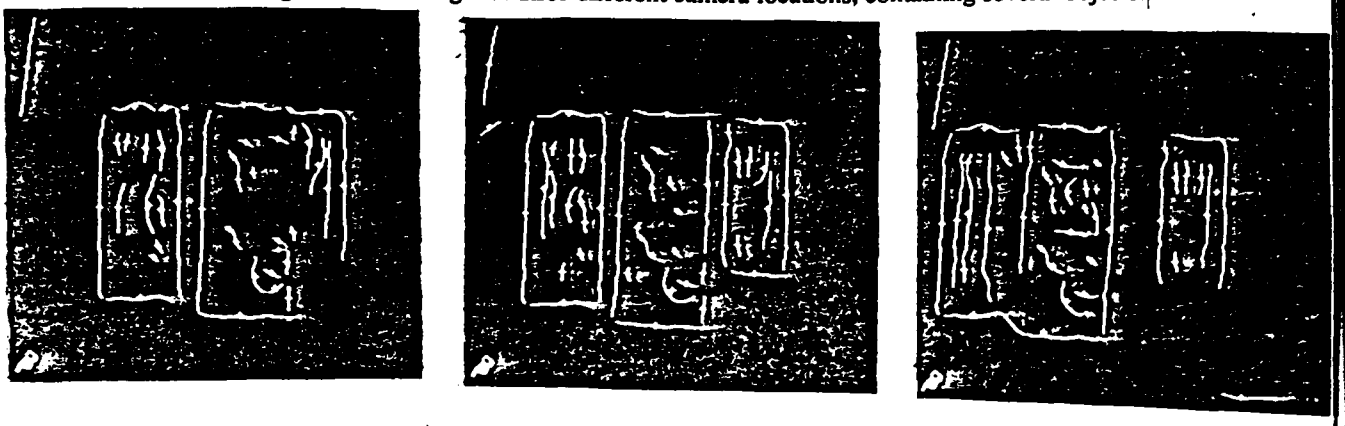
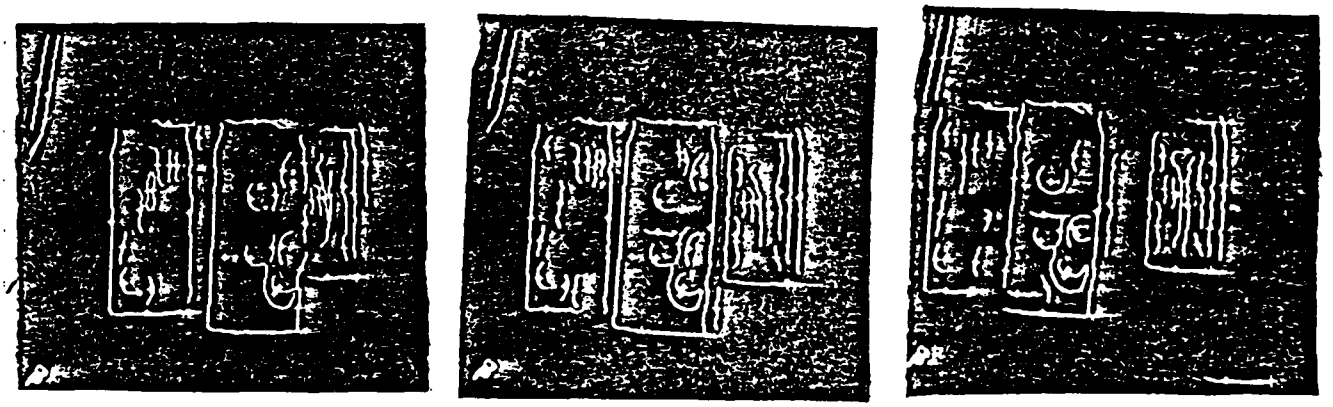


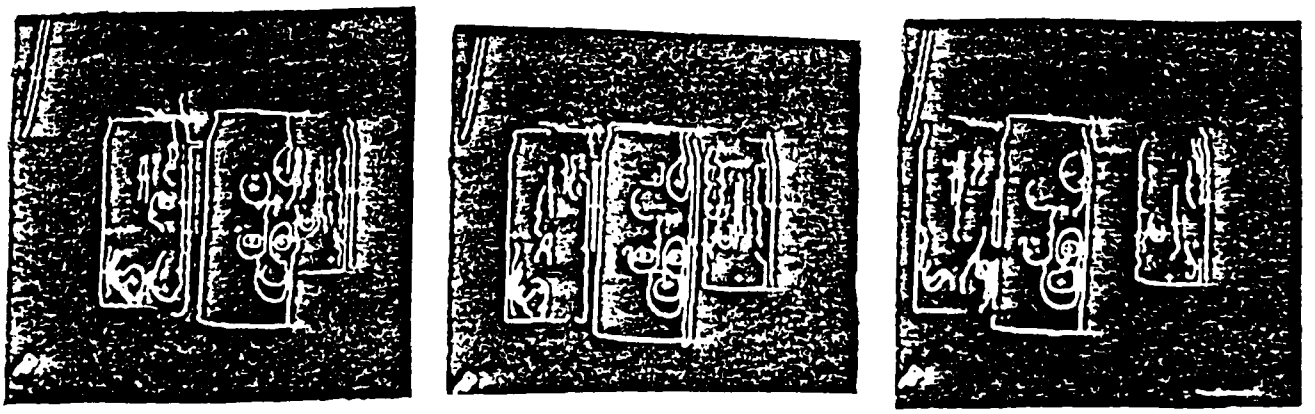
Fig. 9 Original stereo images at three different camera locations, containing several objects.



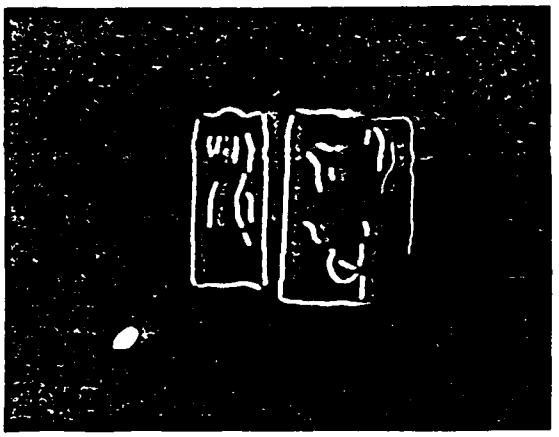
11A) Curve-segments at $\sigma = 6$.



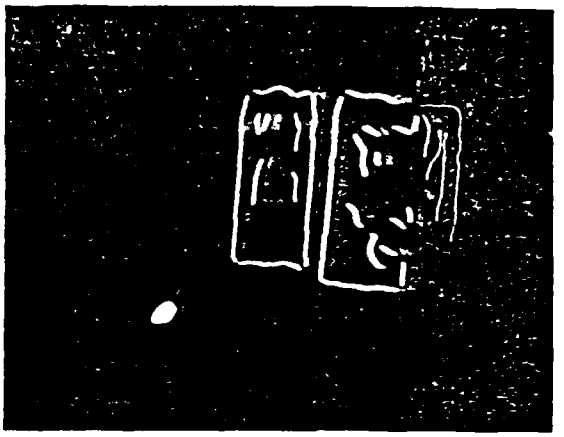
11B) Curve-segments at $\sigma = 3$.



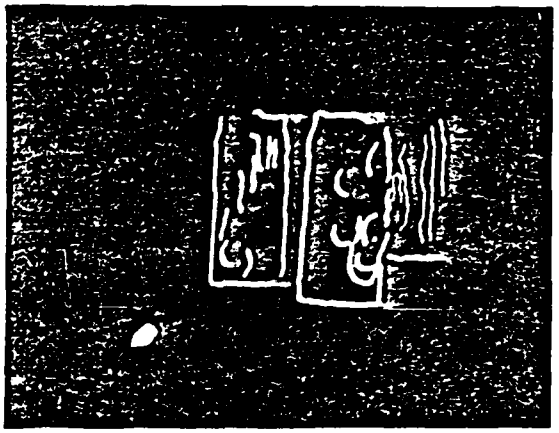
11C) Curve-segments at $\sigma = 1.5$.



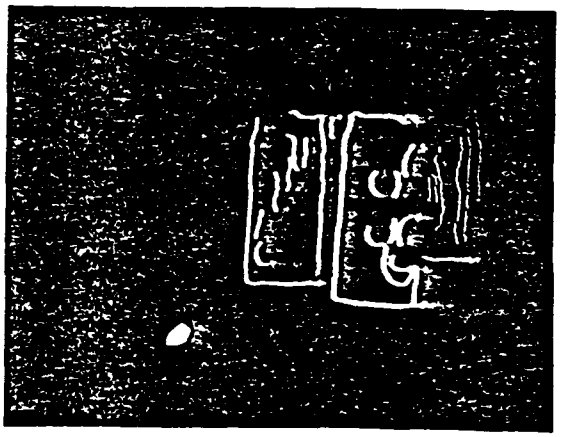
Disparity at σ = 3 between the left and the middle image.



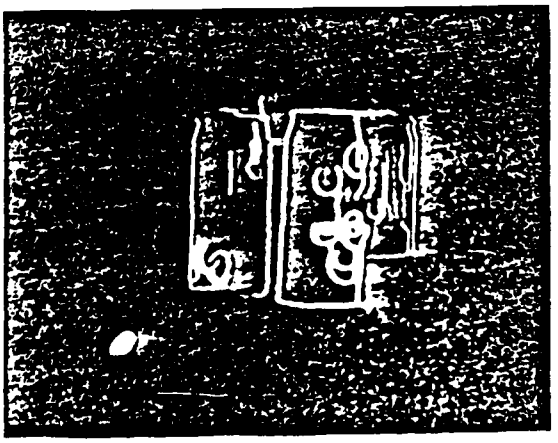
Disparity at σ = 6 between the left and the right image.



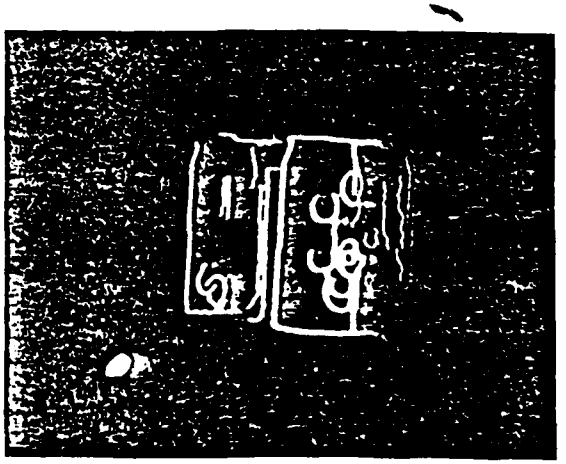
Disparity at σ = 3 between the left and the middle image.



Disparity at σ = 3 between the left and the right image.



Disparity at σ = 1.5 between the left and the middle image.



Disparity at σ = 1.5 between the left and the right image.

Fig. 12 Disparity at σ = 1.5, 3, 6 between the left and the right images as well as the disparity between the left and the right images.

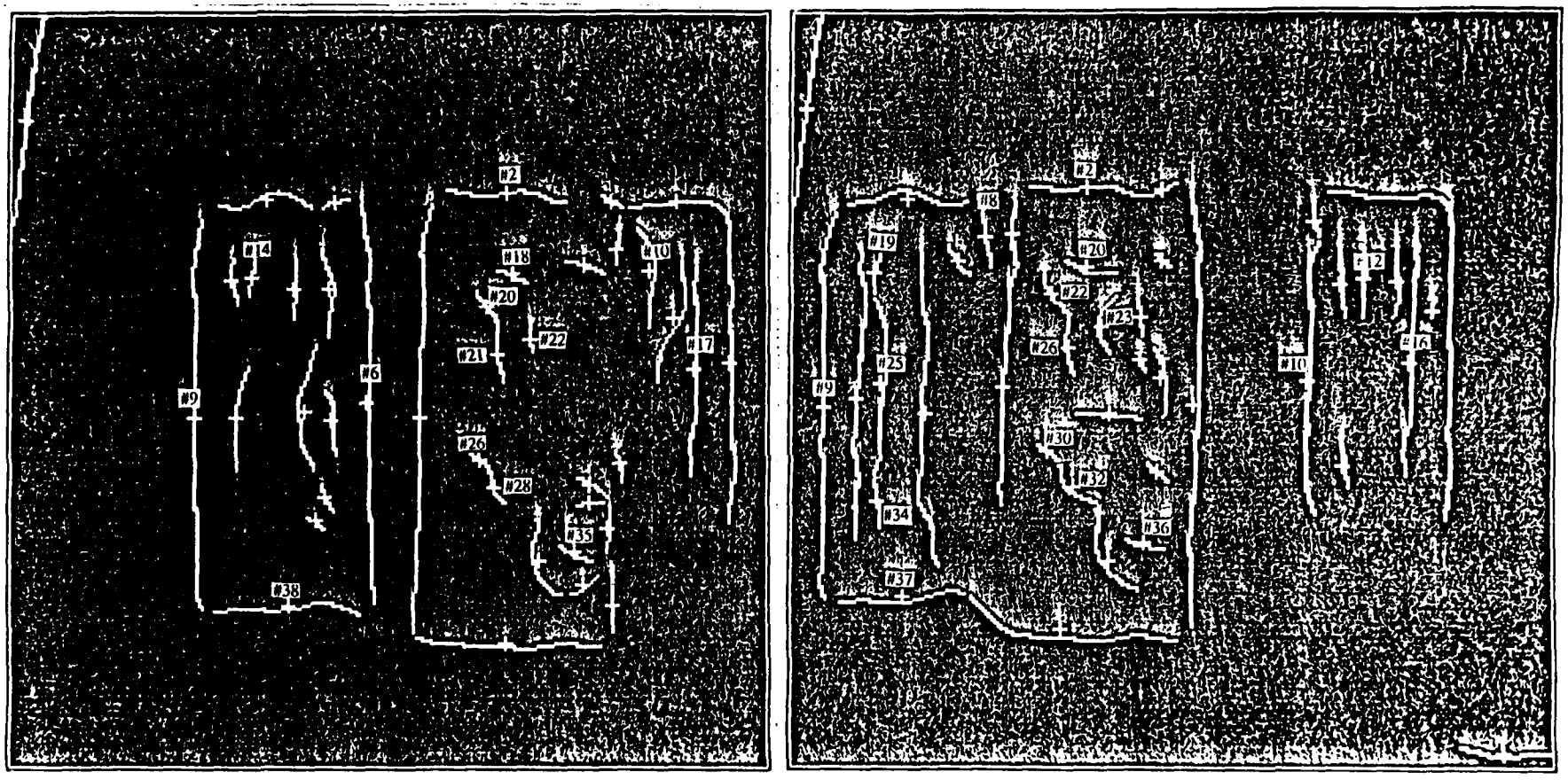


Fig. 13 Extracted curve-segments at $\sigma = 6$, a number of curve-segments are identified and labelled by their centroids locations.

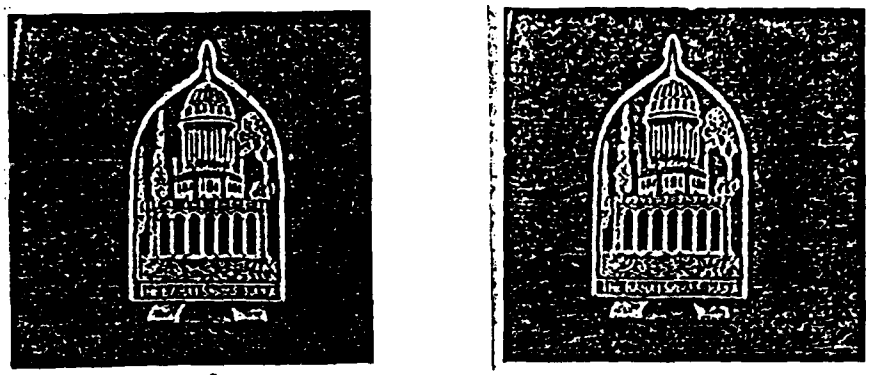
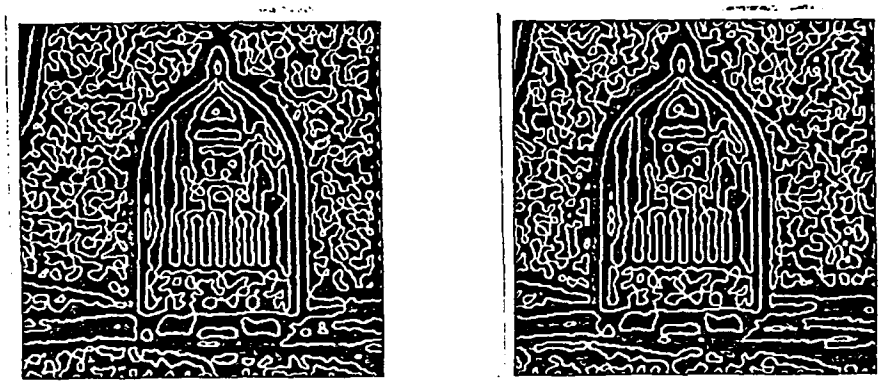
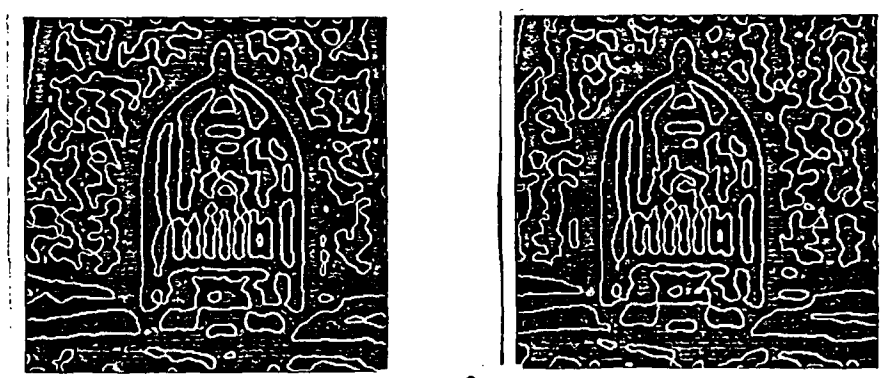


Fig. 14 A pair of stereo images with repetitive patterns of size 256×256.



$\sigma = 1.5$



$\sigma = 3$

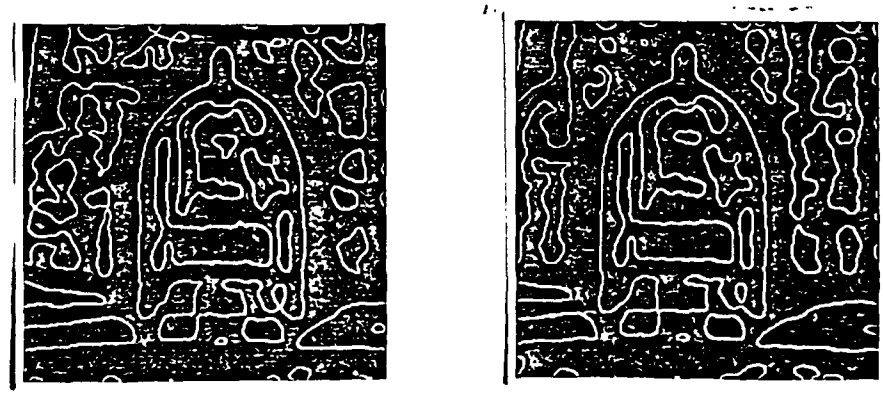
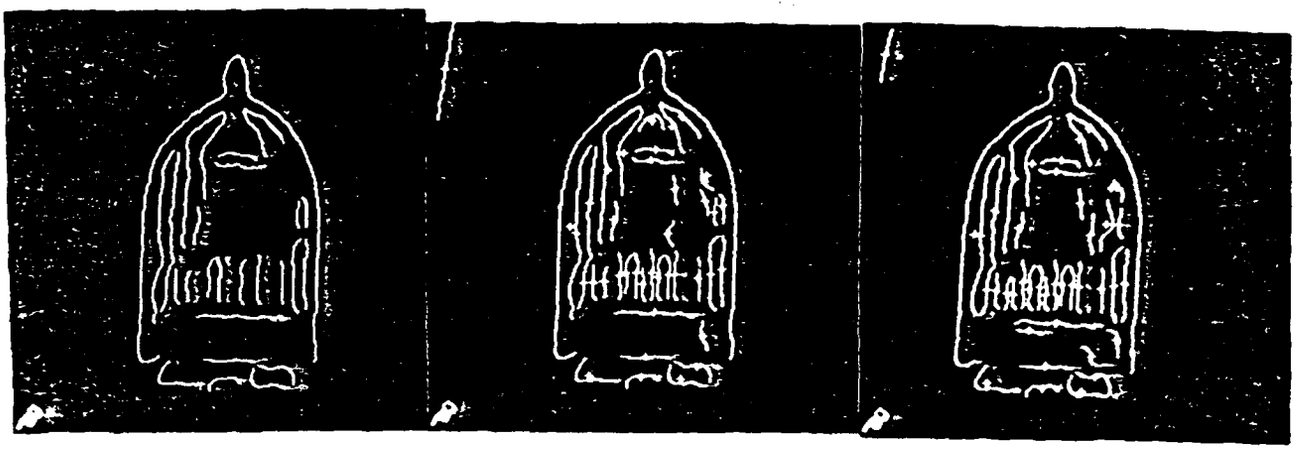
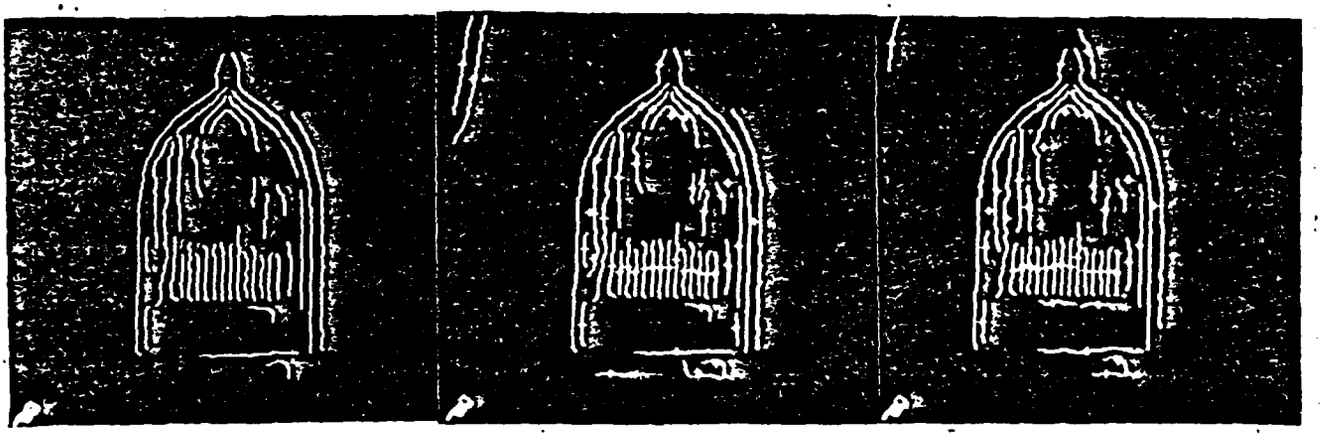


Fig. 15 Zero-crossings at $\sigma = 4.5, 3, 1.5$.



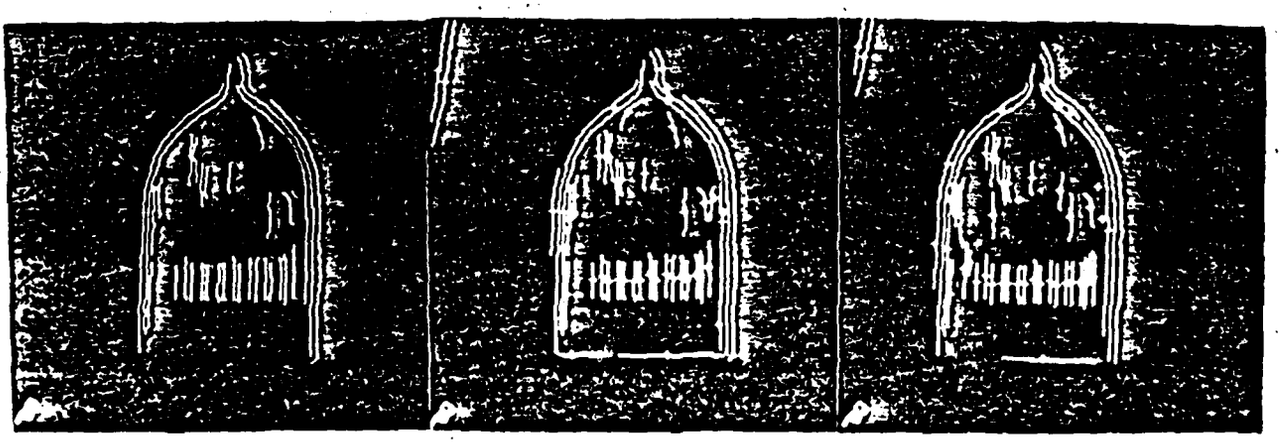
14A) Disparity at $\sigma = 4.5$.

13A) Curve-segments at $\sigma = 4.5$.



14B) Disparity at $\sigma = 3$.

13B) Curve-segments at $\sigma = 3$.



14C) Disparity at $\sigma = 1.5$.

13C) Curve-segments at $\sigma = 1.5$.

Curve-segments and their centroids for $\sigma = 6.0$										
Left Curve-Segments				Right Curve-Segments				Generated Right Curve-Segments		
Curve-#	Y_l	X_l	Curve-length	Curve-#	Y_r	X_r	Curve-length	Curve-#	Y_r	X_r
0	36	5	71	0	33	5	65	40	63	36
1	63	86	39	1	60	124	11	41	59	94
2	61	167	47	2	59	99	43	42	61	36
3	64	202	12	3	63	39	46	43	63	203
4	64	109	10	4	132	135	152	44	134	66
5	64	224	35	5	101	214	155	45	129	74
6	131	120	141	6	75	74	25	46	143	71
7	136	138	146	7	71	174	13	47	118	220
8	118	242	111	8	76	65	23	48	118	202
9	136	62	136	9	133	11	135	49	134	11
10	87	215	40	10	124	173	97	50	134	29
11	80	204	11	11	84	56	18	51	129	31
12	93	107	34	12	90	191	27	52	80	175
13	90	74	18	13	135	45	116	53	85	135
14	89	81	13	14	92	183	25	54	86	74
15	93	95	21	15	91	202	24	55	96	43
16	85	193	18	16	118	207	78	56	87	23
17	120	230	72	17	130	22	102	57	82	11
18	88	169	15	18	84	124	10	58	83	123
19	103	223	42	19	88	28	14	59	103	199
20	99	160	19	20	86	99	22	60	129	44
21	115	164	21	21	126	71	81	61	133	21
22	110	175	10	22	95	87	20	62	136	69
23	134	99	48	23	106	104	24	63	149	136
24	136	76	36	24	103	117	15	64	153	196
25	137	108	27	25	125	30	63	65	155	122
26	151	158	12	26	113	93	22	66	163	70
27	152	205	13	27	123	124	28	67	169	31
28	160	163	17	28	117	117	10	68	170	46
29	159	195	15	29	135	107	25	69	182	135
30	163	106	14	30	148	86	13	70	179	104
31	165	195	10	31	153	185	17	71	199	134
32	174	201	17	32	160	95	26	72	210	98
33	171	103	11	33	155	123	14			
34	185	178	37	34	165	29	19			
35	182	190	17	35	185	107	39			
36	193	193	10	36	179	119	17			
37	200	203	18	37	197	38	45			
38	200	94	55	38	208	91	87			
39	213	167	68	39	250	238	34			

Table 1. It represents the curve-segments in the left and the right images as well as the generated right curve-segments.

Initial Score & Iteratively Relaxed Scores for $\sigma = 6.0$					
Candidate Matching Pairs		Initial, First and Final Relaxation Scores			
left_#	right_#	initial	1st	2nd 4th	final (5th)
1	40	0.8974359	0.0540064		0.00000366
2	2	1.0000000	0.1215373		0.00012471
4	41	0.2000000	0.0050000		0.00000001
4	42	0.8000000	0.0232967		0.00000009
5	43	0.5714286	0.0595047		0.00000324
6	44	0.1560284	0.0039007		0.00000001
7	45	0.1986301	0.0049658		0.00000001
7	46	0.2671233	0.0066781		0.00000001
8	47	0.9639640	0.0618637		0.00000326
8	48	0.1891892	0.0047297		0.00000001
9	49	0.6250000	0.0769739		0.00000570
9	50	0.0955882	0.0106103		0.00000018
9	51	0.1764706	0.0372678		0.00000134
10	12	0.5750000	0.0441748		0.00000217
11	53	0.8181818	0.0218434		0.00001368
12	54	0.1470588	0.0200906		0.00000033
13	55	0.7777778	0.0194445		0.00000001
13	56	1.0000000	0.0872945		0.00000661
14	19	0.6923077	0.0753094		0.00000564
16	58	0.1111111	0.1590943		0.00016946
17	10	0.7500000	0.0259375		0.00000045
17	16	0.8194444	0.0572730		0.00000309
18	20	0.8000000	0.1565279		0.00016516
19	59	0.3333333	0.0555156		0.00000313
20	22	0.8421053	0.1220061		0.00010055
21	26	0.8571429	0.1621467		0.00016728
22	23	1.0000000	0.1642163		0.00016601
24	60	0.7222222	0.0254085		0.00000069
24	61	0.7777778	0.0759626		0.00000573
25	62	0.2222222	0.0055556		0.00000000
26	30	0.9166667	0.1261140		0.00010634
27	63	0.2307692	0.0540056		0.00007407
27	64	0.2307692	0.0255596		0.00000169
28	32	0.8235294	0.0971956		0.00009697
29	65	0.2666667	0.0960830		0.00011155
34	69	0.1891892	0.0182031		0.00003081
34	70	0.8108108	0.0800574		0.00007261
35	36	1.0000000	0.1559567		0.00016836
37	71	0.5000000	0.0455989		0.00006323
38	37	0.7090909	0.0544123		0.00000410
39	72	0.7352941	0.1396286		0.00014726

Table 2. It represents the initial, first and final scores for the relaxation process

Matched curve-segments and their centroids for $\sigma = 6.0$							
Left Curve-Segments			Right Curve-Segments			Centroid	
Curve-#	Y_l	X_l	Curve-#	Y_r	X_r	disparity	
2	61	167	2	59	99	68	Diet Coke (matched)
7	136	138	46	143	71	67	" "
11	80	204	53	85	135	69	" "
16	85	193	58	83	123	70	" "
18	88	169	20	86	99	70	" "
20	99	160	22	95	87	73	" "
21	115	164	26	113	93	71	" "
22	110	175	23	106	104	71	" "
26	151	158	30	148	86	72	" "
27	152	205	63	149	136	69	" "
28	160	163	32	160	95	68	" "
29	159	195	65	155	122	73	" "
34	185	178	70	179	104	74	" "
35	182	190	36	179	119	71	" "
37	200	203	71	199	134	69	" "
39	213	167	72	210	98	69	" "
1	63	86	40	63	36	50	Sunkist (matched)
4	64	109	42	61	36	73	" (mismatch)
6	131	120	44	134	66	54	" "
9	136	62	49	134	11	51	" "
12	93	107	54	86	74	33	" (mismatch)
13	90	74	56	87	23	51	" (matched)
14	89	81	19	88	28	53	" "
24	136	76	61	133	21	55	" "
25	137	108	62	136	69	39	" (mismatch)
38	200	94	37	197	38	56	" (matched)
5	64	224	43	63	203	21	Coca-Cola (matched)
8	118	242	47	118	220	22	" "
10	87	215	12	90	191	24	" "
17	120	230	16	118	207	23	" "
19	103	223	59	103	199	24	" "

Table 3. It represents the matched curve-segments for a pair of stereo images for the coarse channel $\sigma = 6$.

APPENDIX I

CAMERA GEOMETRY

1. Parallel Axis Method

The two cameras are mounted such that their focal axes are parellel and the distance between the two cameras, (baseline), are fixed as shown in Figure 1. Any point in the three dimensional world space, together with the centers of projection of the two camera systems, defines a plane called an epipolar line. In the parallel axis geometry the epipolar lines are parellel to the scan lines, thus the search for finding corresponding points is unidirectional as shown in Figure 1. The point $P(X, Y, Z)$ in the world coordinate system is imaged to point P_r and P_l in the right and left image coordinate plane respectively. The distance $P'_r P_l$, where P'_r is the transformed location of P_r in the left image plane, is known as disparity. It can easily be shown, (see section 3), that the distance, (depth), is inversely proportional to the disparity. Thus, the points which are nearer to the camera will have a larger disparity than points which are further away from the camera.

2. Intersecting Focal Axes

In some situations, the parallel axes method cannot be used because some part of the left image will not be in field of view in the right image due to the lateral shift of the camera. In order for the cameras to have the same field of view, they are rotated about their y-axes such that their focal axes intersect at a point. This point is known as the fixation point as shown in Figure 2 [2 pp. 303].

A point on the object will cast image points A and B in the left and the right cameras. Associated with each image element is its angular displacement from the optic axis as shown in Figure 3. If α_1 and α_2 are the angular displacements to the two image elements corresponding to the same object point, the disparity, d, of the object

point is defined as $d = \alpha_1 + \alpha_2$. The objects in front of the fixation point will have positive disparities, (convergent), and objects behind the fixation point will have a negative disparities, (divergent), as shown in Figure 3 and 4 respectively. Figure 5 shows a stereo pair where the fixation point is somewhere in the middle of the scene. Looking at the right image, the blocks in front of the fixation point possess positive disparities, and the blocks behind the fixation point possess negative disparities compared to the left image.

Disparity values give a depth measure relative to the fixation point since the fixation point has zero disparity. In order to obtain the absolute depth values, the camera's orientation with respect to a fixed coordinate system has to be known as well as the distance between the two cameras. The absolute depth is usually measured with respect to the coordinate system of one camera.

3. CAMERA MODELLING

In order to obtain the relative orientation of the cameras, we have to know how the image is formed in the first plane. Mapping of the world coordinate points into the image plane is known as the perspective transformation. Consider that the camera can be represented by a pinhole, thus, the distortion due to the lens can be ignored, and its position is measured with respect to a fixed coordinate system known as the world coordinate. Let the Cartesian components of vector $\vec{C}$ represent the coordinates of the focal center as shown in Figure 6. Orientation of the camera is denoted by the unit vector $\hat{a}$, which is perpendicular to the image plane. We will use two more unit vectors $\hat{H}$ and $\hat{V}$, orthogonal to each other, and that are also orthogonal to the aiming unit vector $\hat{a}$. In terms of $\hat{H}$, $\hat{V}$, and $\hat{a}$ all the points in the image plane are described by $(\vec{C} - f\hat{a} + x'H + y'V)$ for different values of the scalar parameters u and v.

The ordered pair (x', y') could be considered to be the coordinates of a point in the image plane. The perspective transformation equations will relate image coordinates (x', y') to the object point $\vec{P}$. Comparing the appropriate similar triangles we obtain the following equations

$$\frac{x'}{f} = \frac{\vec{D}\cdot\hat{H}}{\vec{D}\cdot\hat{a}} \tag{1a}$$

$$\frac{y'}{f} = \frac{\vec{D}\cdot\hat{V}}{\vec{D}\cdot\hat{a}} \tag{1b}$$

where $\vec{D} = \vec{P} - \vec{C}$ as shown in Figure 6, is the vector from the camera focal center to the object point. Its component along the axes is obtainable by taking the scalar product about each axes, thus, the physical point $P(X, Y, Z) = P(\vec{D}\cdot\hat{H}, \vec{D}\cdot\hat{V}, \vec{D}\cdot\hat{a})$.

Consider the parallel axes camera geometry where optical axes are parallel to one another and perpendicular to the baseline connecting the two cameras as shown in Figure 7. Let the world coordinate system (X, Y, Z) be placed midway between the lens centers, and the image coordinates in the left and right image be (x'_l, y'_l) and (x'_r, y'_r) respectively. Then

$$\frac{x'_l}{f} = \frac{X + \frac{b}{2}}{Z} \quad \text{and} \quad \frac{x'_r}{f} = \frac{X - \frac{b}{2}}{Z} \quad \textit{while} \quad \frac{y'_l}{f} = \frac{y'_r}{f} = \frac{Y}{Z}\,. \tag{2}$$

where f is the distance from the lens center to the image plane in both cameras and b is the distance between the lens centers. Using the above three relationships we can solve for the three unknowns x, y, and z.

$$X = b\,\frac{(x'_l + x'_r)/2}{x'_l - x'_r} \tag{3}$$

$$Y = b\,\frac{(y'_l + y'_r)/2}{x'_l - x'_r} \tag{4}$$

$$Z = b\,\frac{f}{x'_l - x'_r} \tag{5}$$

where the difference in image coordinates $(x'_l - x'_r)$ is the disparity. From equation 5 it is seen that the accuracy of Z measurement depends on the baseline and disparity measurements. A larger baseline will give rise to a larger disparity and thus a better Z measurement, but the disparity measurement is less accurate because obtaining the corresponding points x'_l, x'_r is now more difficult.

Consider now that the cameras are set up such that their optical axes intersect at a fixation point. In order to obtain the three dimensional coordinates of the object points we have to know the relative orientation of the cameras with respect to the world coordinate.

Figure 8 [2 pp. 303] represents the coordinates of the two camera systems. The transformation from one camera station to another can be represented by a translation and a rotation. Thus, if $r_l = (x_l, y_l, z_l)^T$ is the position of a point P in the left camera coordinate system and $r_r = (x_r, y_r, z_r)$ is the position of the same point P in the right camera coordinate system, then

$$r_r = Rr_l + T \tag{6}$$

where R is a 3 x 3 orthogonal matrix representing the rotation, while T is a vector corresponding to the translation.

The rotation matrix can be decomposed into three components, a rotation about x-axis (tilting), a rotation about y-axis (panning), and a rotation about z-axis (rolling) as shown in Figure 9 and 10,

$$R = R_x R_y R_z \tag{7}$$

where

$$R_x = \begin{bmatrix} cos(\beta 1) & 0 & sin(\beta 1) \\ 0 & 1 & 0 \\ -sin(\beta 1) & 0 & cos(\beta 1) \end{bmatrix} \tag{7a}$$

$$R_y = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\beta 2) & sin(\beta 2) \\ 0 & -sin(\beta 2) & cos(\beta 2) \end{bmatrix} \tag{7b}$$

$$R_z = \begin{bmatrix} cos(\beta 3) & sin(\beta 3) & 0 \\ -sin(\beta 3) & cos(\beta 3) & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{7c}$$

where $\beta 1$, $\beta 2$, $\beta 3$ are the corresponding rotation angles about x-axis, y-axis, and z-axis respectively.

T is a vector representing the distance between the two cameras which can be decompared into

$$T = b\ T_{a1}\ T_{a2} \tag{8}$$

where b is a constant and T_{a1} , and T_{a2} represents the direction of the vector T

$$T_{a1} = \begin{bmatrix} cos(\alpha 1) & 0 & sin(\alpha 1) \\ 0 & 1 & 0 \\ -sin(\alpha 1) & 0 & cos(\alpha 1) \end{bmatrix} \tag{8a}$$

$$T_{\alpha 2} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos(\alpha 2) & sin(\alpha 2) \\ 0 & -sin(\alpha 2) & cos(\alpha 2) \end{bmatrix} \tag{8b}$$

α_1, and α_2 are elevation and azimuth angles respectively.

In order to use the stereo system to obtain the absolute distance, the parameters α_1, α_2, β_1, β_2, β_3 and the focal length of the cameras have to be known. To obtain these parameters, one takes a set of points where 3-D coordinates are known and a least-square minimization technique is used to solve for the parameters.

In this report, we assume that the camera's paremeters are known, and then the problem is to obtain the absolute distances or disparities. Once R and T are known, one can compute the position of a point with known left and right image coordinates. If (x'_l, y'_l) and (x'_r, y'_r) are these image coordinates, then from Equation 6 we have

$$(r_{11} \frac{x'_l}{f} + r_{12} \frac{y'_l}{f} + r_{13}) Z_l + r_{14} = \frac{x'_r}{f} Z_r \tag{9a}$$

$$(r_{21} \frac{x'_l}{f} + r_{22} \frac{y'_l}{f} + r_{23}) Z_l + r_{24} = \frac{y'_r}{f} Z_r \tag{9b}$$

$$(r_{31} \frac{x'_l}{f} + r_{32} \frac{y'_l}{f} + r_{33}) Z_l + r_{34} = Z_r \tag{9c}$$

where

$$R = \begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix}, \quad T = \begin{bmatrix} r_{14} \\ r_{24} \\ r_{34} \end{bmatrix}$$

and

$$\frac{x'_r}{f} = \frac{x_r}{Z_r}, \quad \frac{y'_r}{f} = \frac{y_r}{Z_r}$$

We can use any two of the above equations to solve for Z_l and Z_r.

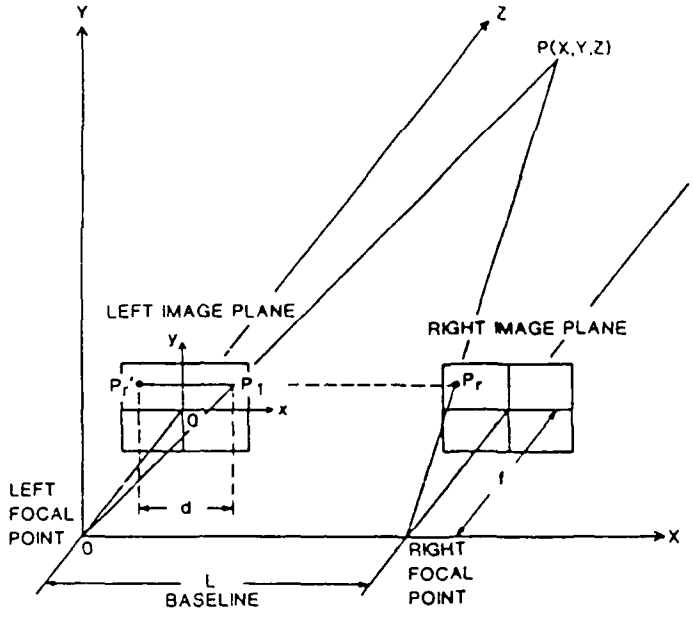


Fig. 1 Parallel axes method, the cameras are set up such that their focal axes are parallel and the line joining the focal centers is perpendicular to it [22].

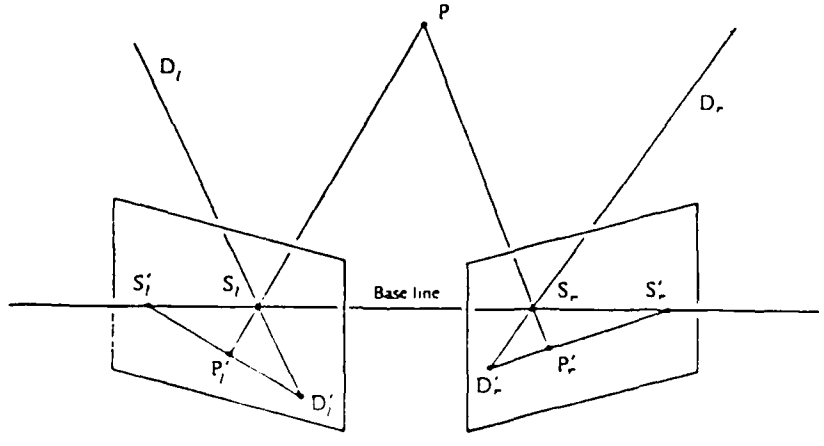


Fig. 2 The images, P'_l and P'_r, of a point P in the environment must lie on corresponding epipolar lines [2 PP. 312].

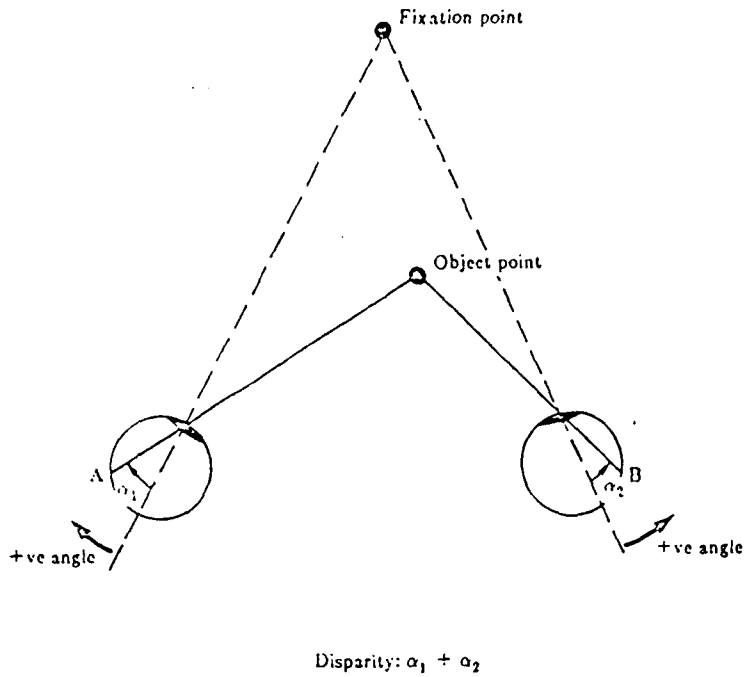


Fig. 3 An object point possesses convergent disparity when it lies in front of the fixation point [59].

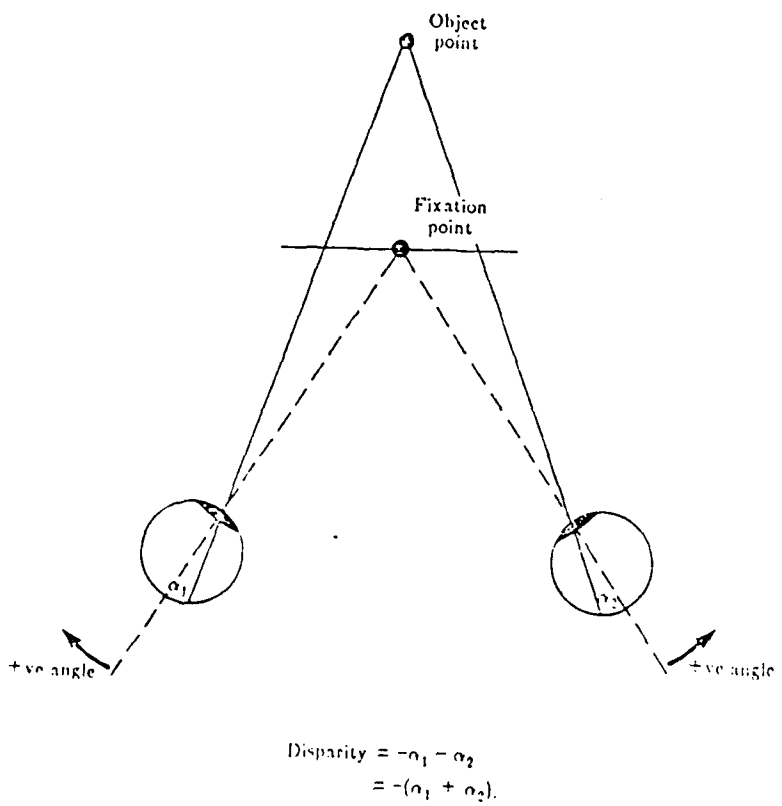
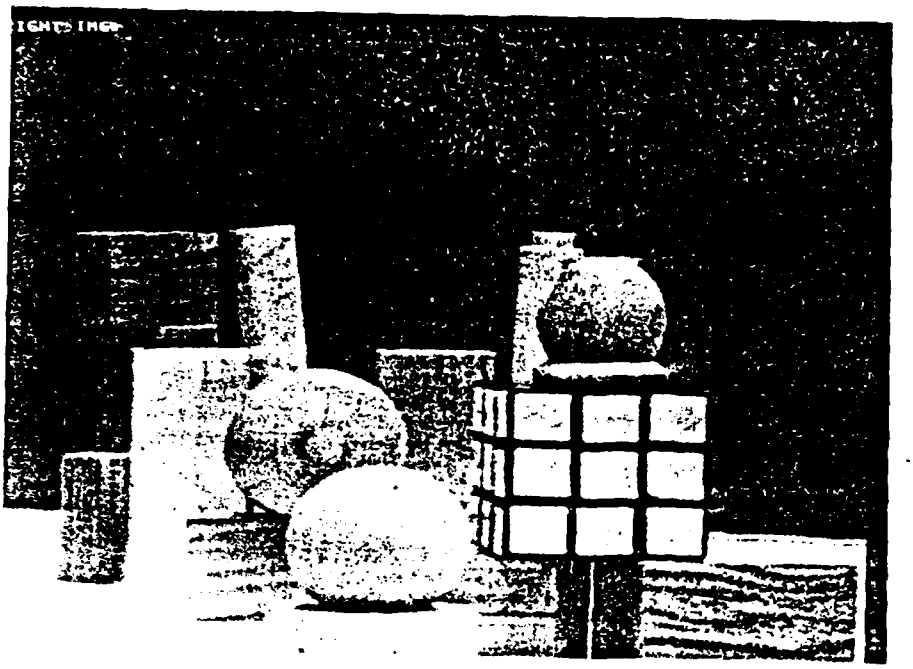
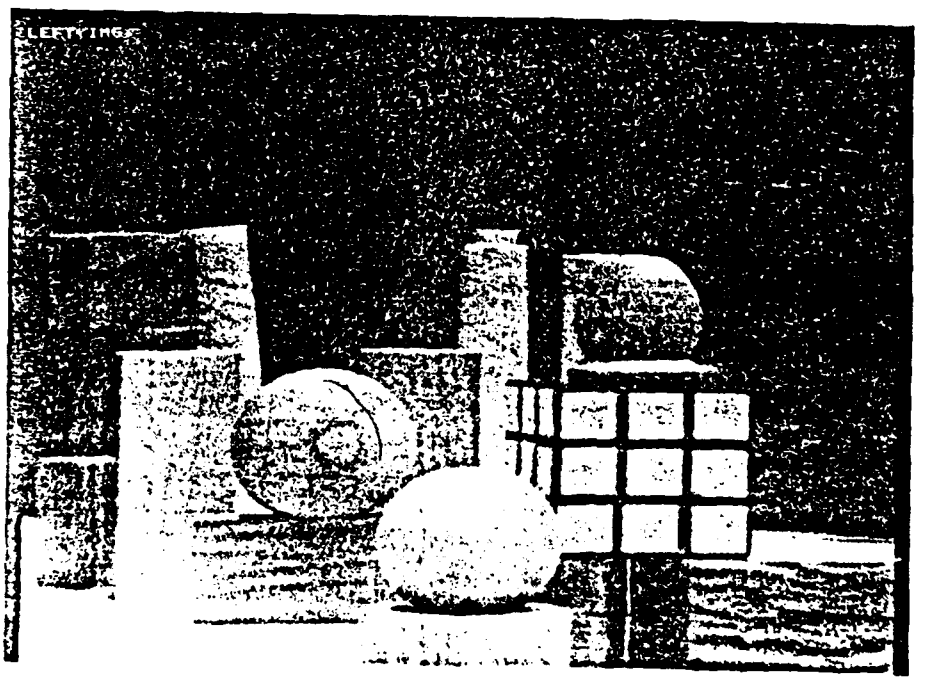


Fig. 4 An object point possesses divergent disparity when it lies in behind of the fixation point [59].



right image



left image

Fig. 5 Stereo pair of a block scene with the fixation point some where in the middle of the scene.

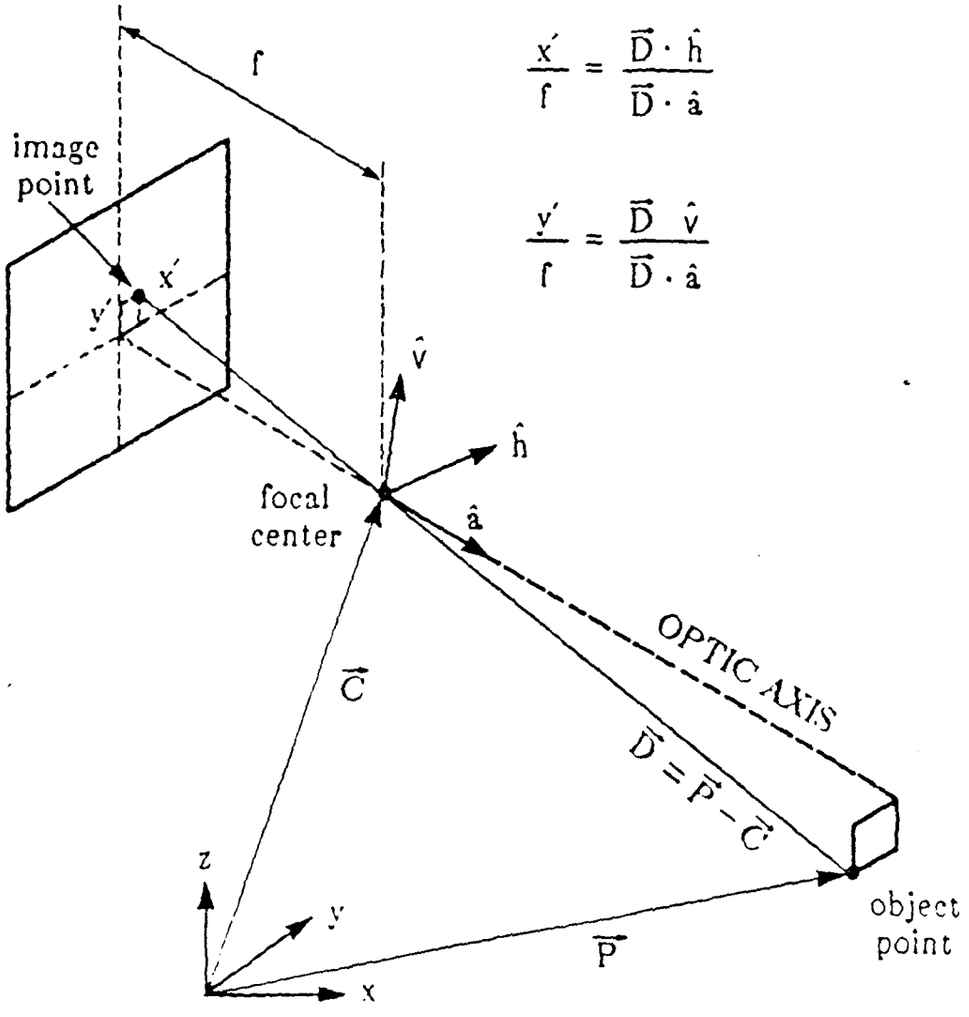


Fig. 6 Image formation, the world coordinate is denoted by (X,Y,Z) [59].

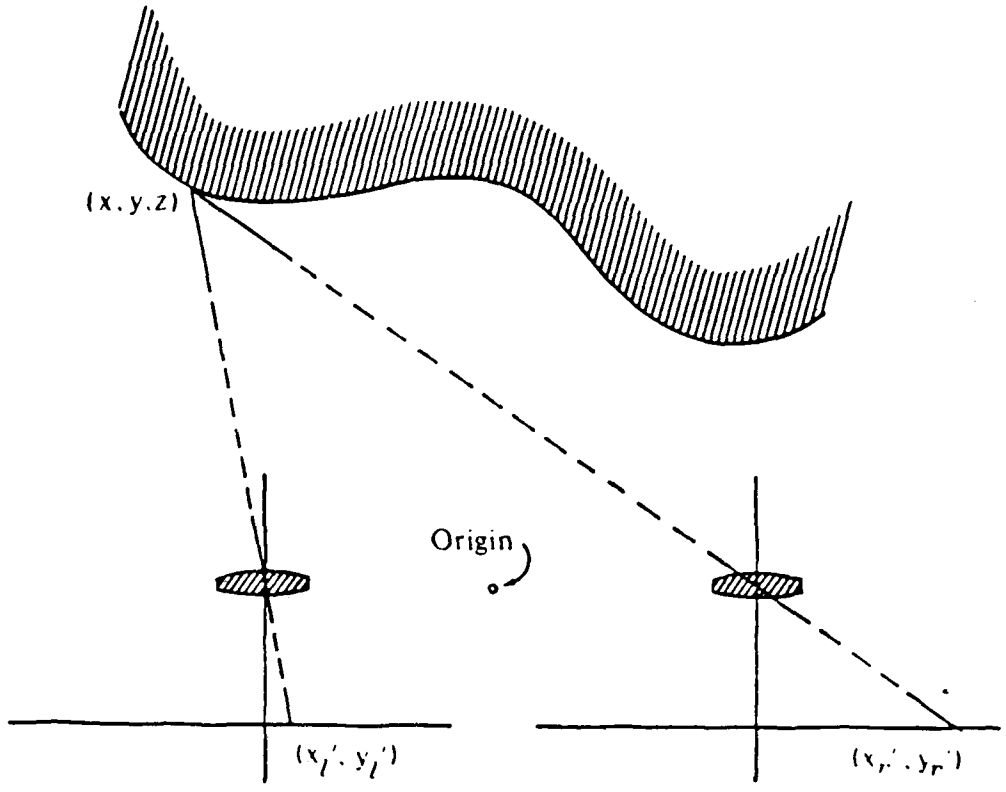


Fig. 7 Parallel axes geometry, the world coordinate origin is midway between the lens centers [2 PP. 300].

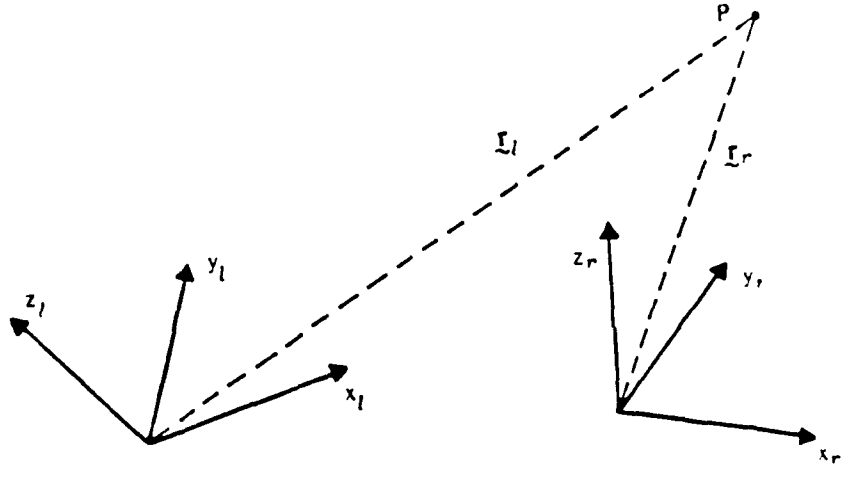


Fig. 8 The relationship between the coordinates, r_l and r_r , of a point P can be given by means of a offset vector and a rotation matrix [2 PP. 303].

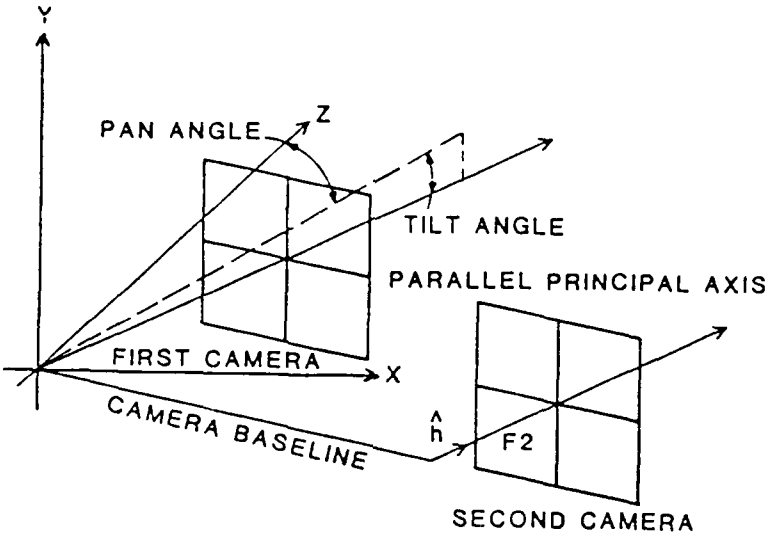


Fig. 9 Shows the pan and tilt angles.

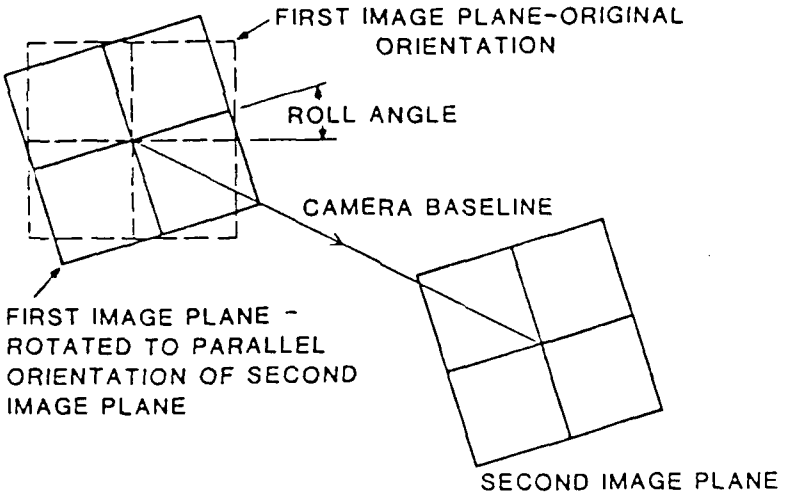


Fig. 10 Shows the rolling angle.

APPENDIX II

Marr-Poggio-Grimson Stereo Algorithm

1. Introduction

The basic problem, as mentioned before, in binocular fusion is the correspondence problem. This is because of the abundance of matchable features in the stereo pair image and the disparity range over which matches are sought. The Marr-Poggio-Grimson algorithm [29] is a multi-channel stereo technique. They suggested the idea of matching coarse, widely separated features first, and then with the information so obtained, repeat the matching process at successively finer scales of resolution. Coarse channels are expected to control the vergence movements, that is to cause fine channels to come into correspondence. The coarse-to-fine strategy will decrease the false targets as well as reduce the searching time. This algorithm has been simulated by the principal investigator and results are presented in the following subsections.

Marr-Poggio suggested the use of zero-crossings, labelled by the sign of their contrast change and their rough orientation in the image as good candidate points for matching.

2. Marr-Hilderth Edge Operator

Marr-Hilderth suggested the use of a orientation-independent operator combined with a Gaussian filter. They chose the Lapacian operator as given below:

$$\nabla^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \tag{1}$$

which is equivalent to the mask

$$\nabla^2 = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix} \tag{2}$$

The gaussian filter is given by

$$G(x,y) = \sigma^2 \exp\left\{\frac{-(i^2 + j^2)}{\sigma^2}\right\} \tag{3}$$

Figure 1 shows an ideal edge. The first derivative will give rise to a peak and the second derivative to a zero-crossing. This zero-crossing, the place where the value changes from positive to negative is the location of the edge. First images have to be Gaussian filtered then the effect due to noise is reduced. It is also possible to detect coarse and then fine edges by using different filter size.

We can combine the Lapacian filter with the Gaussian filter by obtaining the second derivative of Equation 3, and with respect to x and y we get

$$\nabla^2 G(i,j) = \left[\frac{(i^2 + j^2)}{\sigma^2} - 2\right] \exp\left\{\frac{-(i^2 + j^2)}{2\sigma^2}\right\} \tag{4}$$

Figure 2 shows the impulse response of this filter for different values of the constant which also determines the size of the filter.

The operator is really a band-pass filter which can also be shown to approximate the difference of Gaussians (DOG) [35]. This operator has several important features. It is oriented invariant, thus, there is no need for several edge masks as in most edge detectors. It is possible to operate at different bands, thus large filters can be used to detect blurry shadow edges, and small ones to detect sharply focussed fine detail in the image. Figure 3 and 4 show for example the edge detected images at different $\sigma = 1.5$, 3, and 6 for two SEM images. It is clear that for feature extraction and interpretation several channels are required. The stereo algorithm discussed in the next section is based on this multi-channel edge detector.

3. Matching Process

One important parameter in multi-channel stereo algorithms is the size of the channel. From equation 5, the primal sketch operator, it is seen that the distance between the first zeroes on either side of the origin is given by

$$W_{2D} = 2\sqrt{2\sigma} \tag{5}$$

By convolving the image with a primal sketch operator or width W_{2D}, we have ensured that most zero-crossing contours will be at least W_{2D} apart from one another. Thus the maximum disparity, d, should be equal or less than the largest filter in the multi-channel stereo technique.

The matching process is assumed one dimenionsal, thus, images are assumed to have no or very little vertical disparities. All the candidate points in the search neighborhood that represent the same sign change, and the same zero-crossing-orientations are considered to be potential matches. Of all the potential matches, we assume that only one is a true match. All others will be considered to be false targets. Occlusion is also ignored.

The human visual system is known to possess five different channels for disparity calculation. The values of W_{2D} for these channels are approximately 63, 35, 17, 9, and 4 pixels. Here a pixel here is meant to be the size of a foveal receptor, one such receptor corresponds roughly to an angular interval of 0.4 of the arc. Therefore, if we digitize a visual angle of 4 on the side into a 650 x 650 matrix, we will match the sampling capability of the foveal of the human eye.

4. Implementation

The primal sketch operator of the largest size is applied on the stereo image. The location as well as the contrast change of each pixel is stored in a buffer for each image. The orientation for each pixel is obtained by estimating the δx and δy components of the local gray level gradient

$$\delta x = (A_3 + 2A_4 + A_5) - (A_1 + 2A_8 + A_7) \tag{6}$$

and

$$\delta y = (A_2 + 2A_2 + A_3) - (A_7 + 2A_6 + A_5) \tag{7}$$

where A_m's are the eight neighboring pixels surrounding the central pixel as shown below

$$\begin{bmatrix} A_1 & A_2 & A_3 \\ A_8 & f(x, y) & A_4 \\ A_7 & A_6 & A_5 \end{bmatrix} \tag{8}$$

The direction, θ, of the gradient is then determined from the following relationship,

$$\theta = \tan^{-1} \frac{\delta y}{\delta x} \tag{9}$$

This angle is then classified into six intervals of 30 . Given the location of a zero-crossing and its attribution in the left image we will limit the research neighborhood to an interval of W_{2D} on either side of the same location of the right image. If there is only one potential match, then that match is accepted and the disparity associated with this match is computed and assigned to the candidate point in question. If there is more than one potential match, the following procedure is used for disambiguity between them. All the potential matches within the search neighborhood are divided into three pools. These pools consist of two larger convergent and divergent regions and a smaller one lying centrally between them. If more than one potential match is found in any of the three pools, then no match is assigned to the candidate point. If more than one pool contains a potential match, then the candidate pixel is considered to have ambiguous matches. The ambiguity is resolved by using what is known as the pulling effect, (continuity), which consists of examining the unambiguous disparities within the neighborhood of the candidate point in question of the potential matches available by choosing one that is dominant within the neighborhood.

The above matching process is then repeated for the finer channel. The disparity values from the coarser channel is used to bring image regions within the range of fusion of the finer channel. This is done on a region by region base, an average disparity is evaluated around each candidate point using coarser channel disparity values. In human visual systems this is equivalent to the change in the fixation point.

5. Results Obtained with a Three Channel Stereo Algorithm

Simulation of the stereo algorithm using three channels is discussed. The constant σ = 1.5, 3. and 6 pixels will provide the primal sketch operation of width W_{2D} = 4, 8, and 17 pixels.

Two stereo pair images are displayed in Figure 5. The output of the primal sketch operator for σ = 1.5, 3, and 6 for the two stereo image pairs are shown in

Figure 6 and 7. Note the sign of the contrast change of the edge is gray level coded to gray or white.

Figure 8 and 9 shows the orientation using equation 9. The values are gray level coded.

The disparity values are shown in Figure 10 and 11. The output of each channel is also shown. Consider the disparity of the paperwad stereo. The points which are nearer to the camera have larger disparities than the points which are further away. The paperwad stereo image is a very good example since it has information at different frequencies. As seen from Figure 11, the disparities of coarse features are first obtained and then fine features are matched.

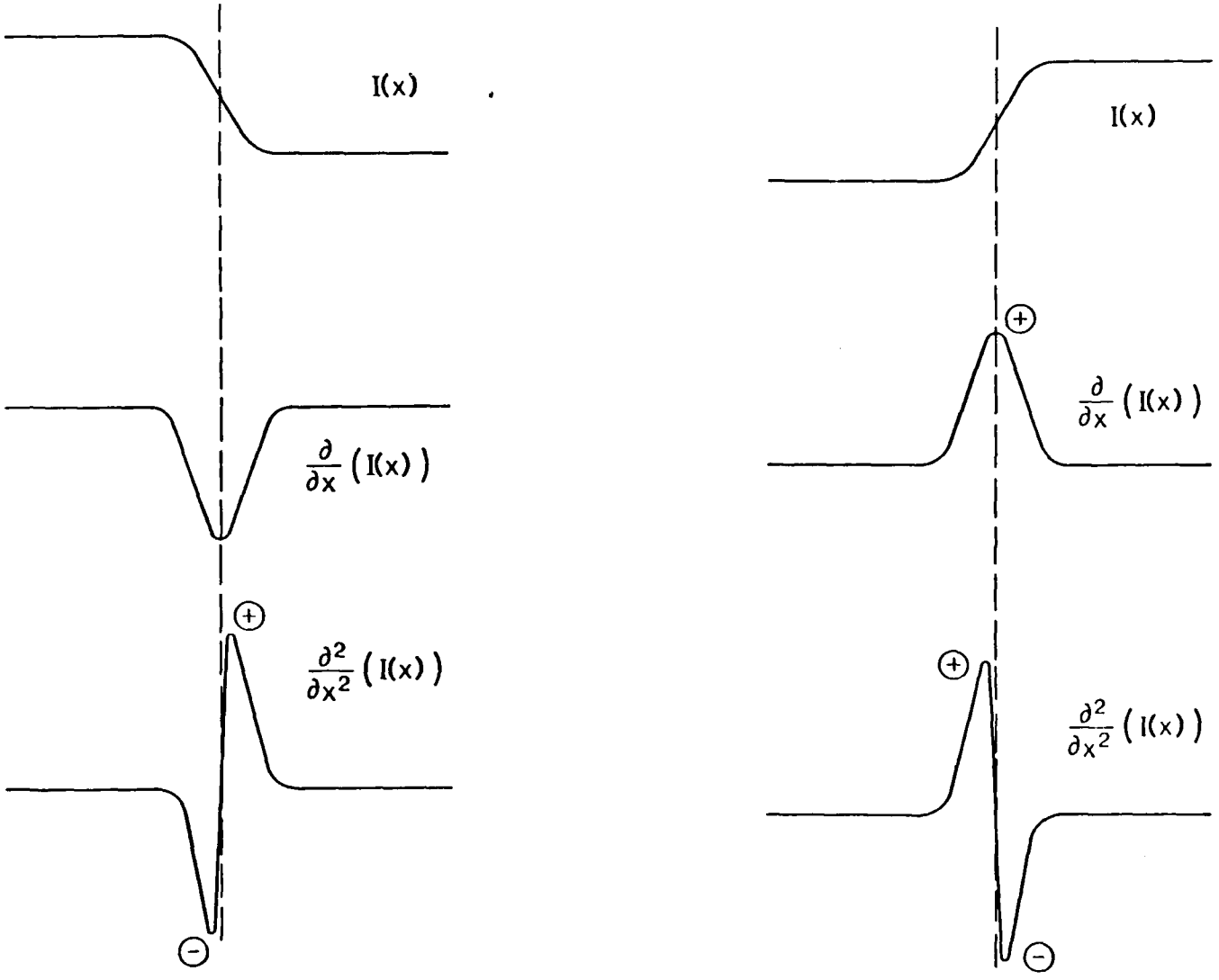


Fig. 1 Shows an ideal edge and its first and second derivative respectively, the two cases of a step up and down edge is shown.

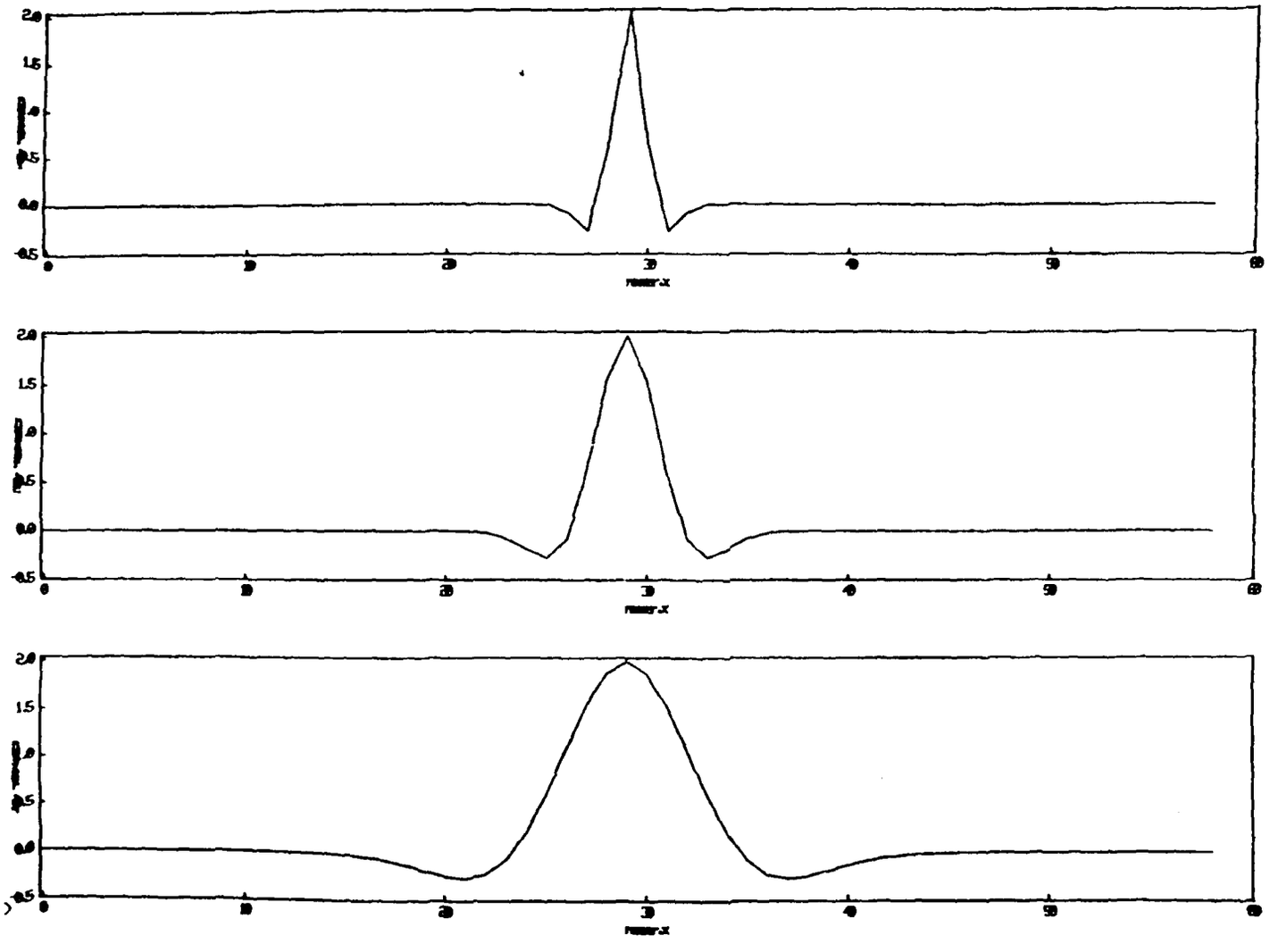


Fig. 2 Output response of the primal sketch operator for different values of σ = 1.5, 3, and 6.

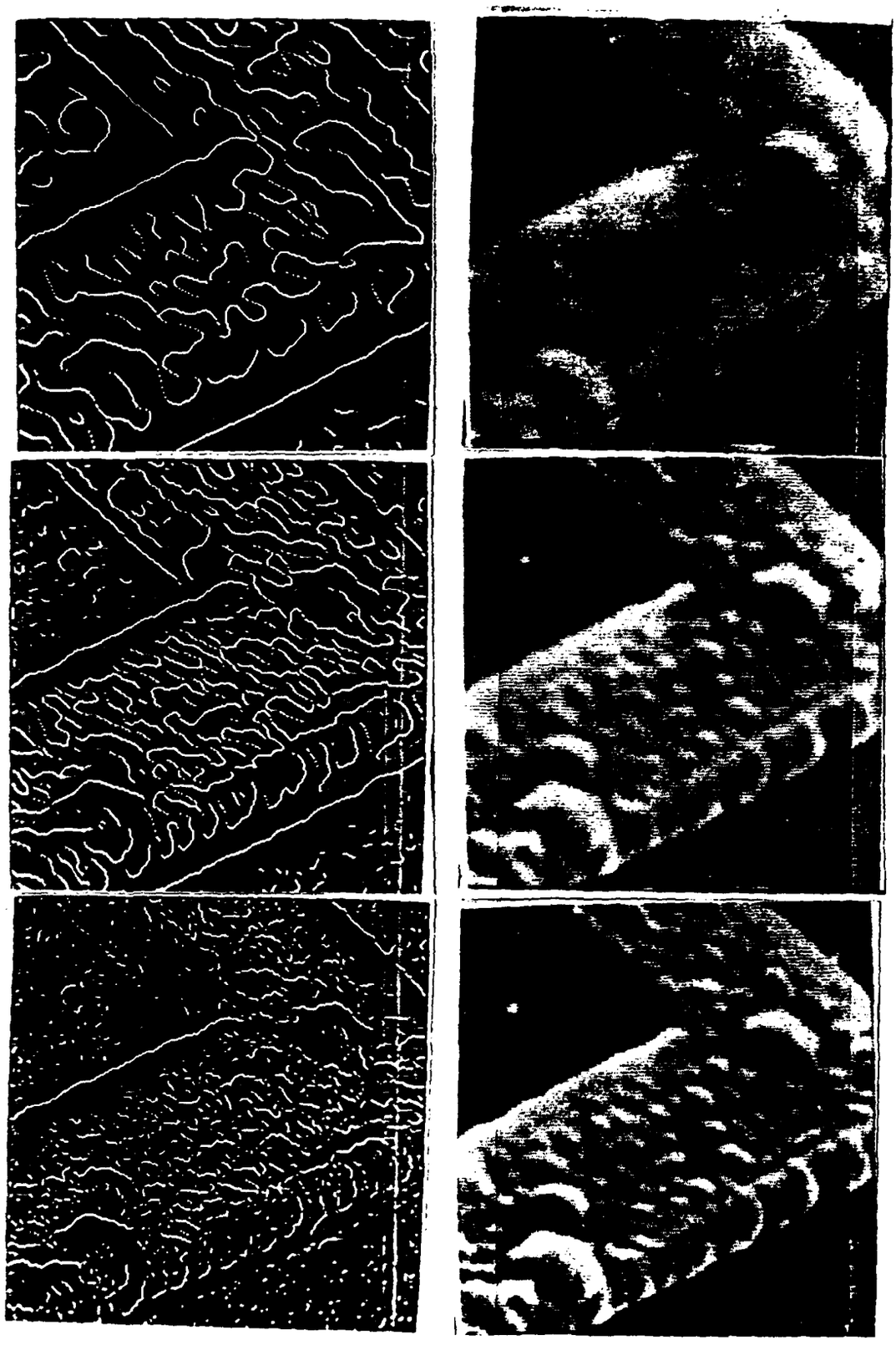


Fig. 3 Zero-Crossings of an SEM image of a VLSI chip, for σ = 1.5, 3, and 6, a threshold of T=50 [60].

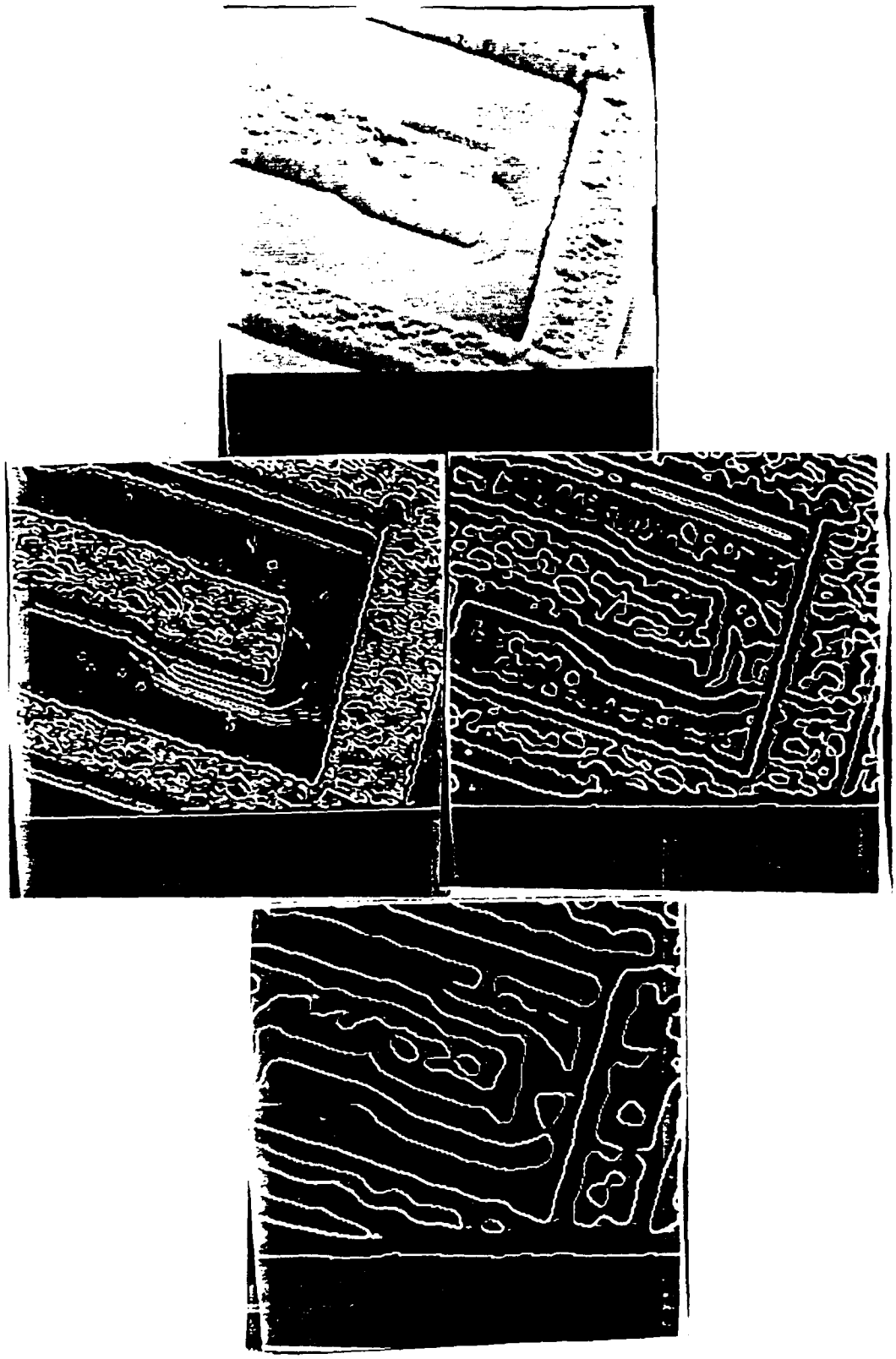


Fig. 4 Zero-Crossings of an SEM image of a VLSI chip, for σ = 1.5, 3, and 6, a threshold of T=60 [60].

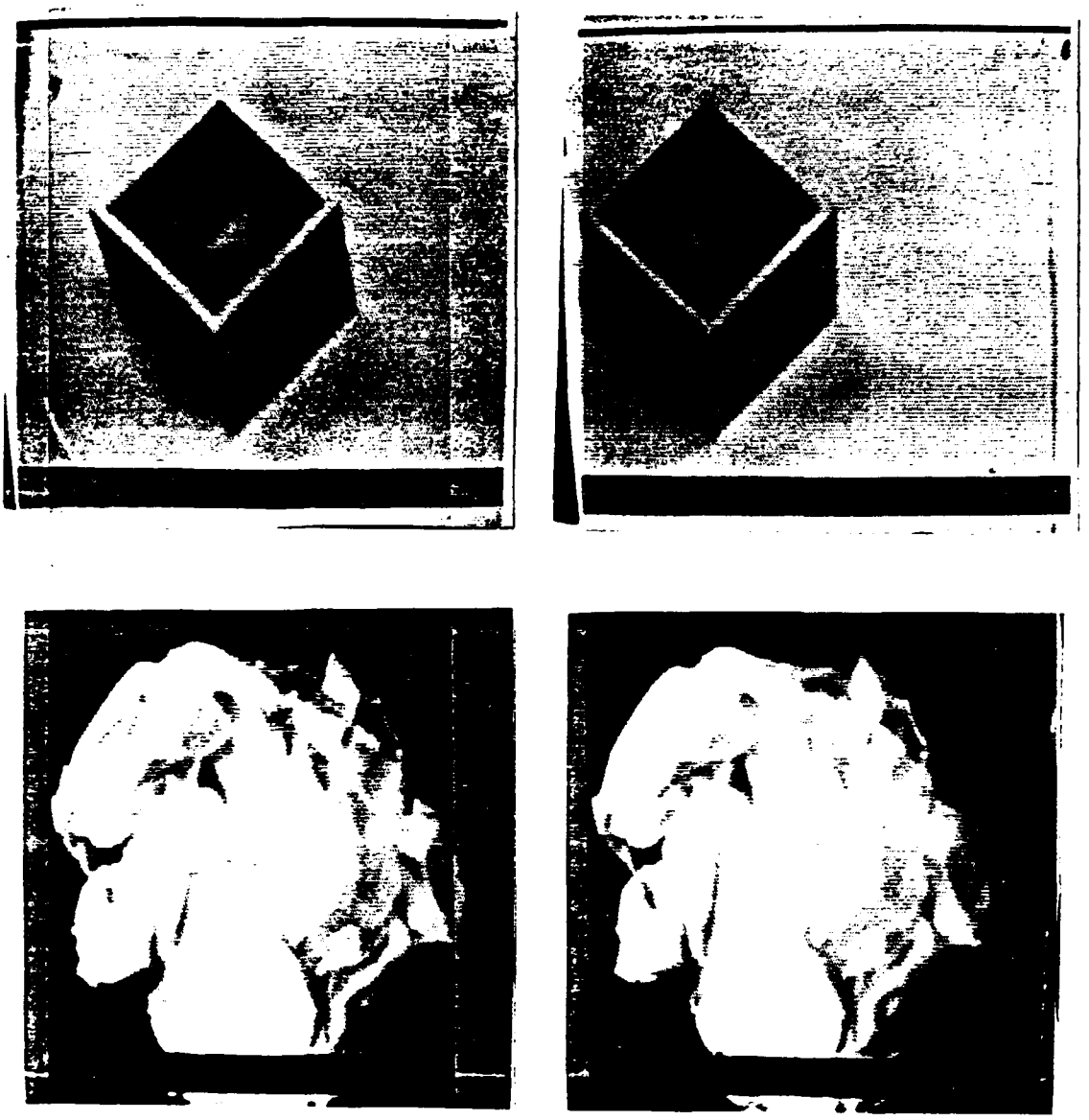


Fig. 5 Shows stereo images of a box and paper-wad, images are of size 256*256 and 8 bits.

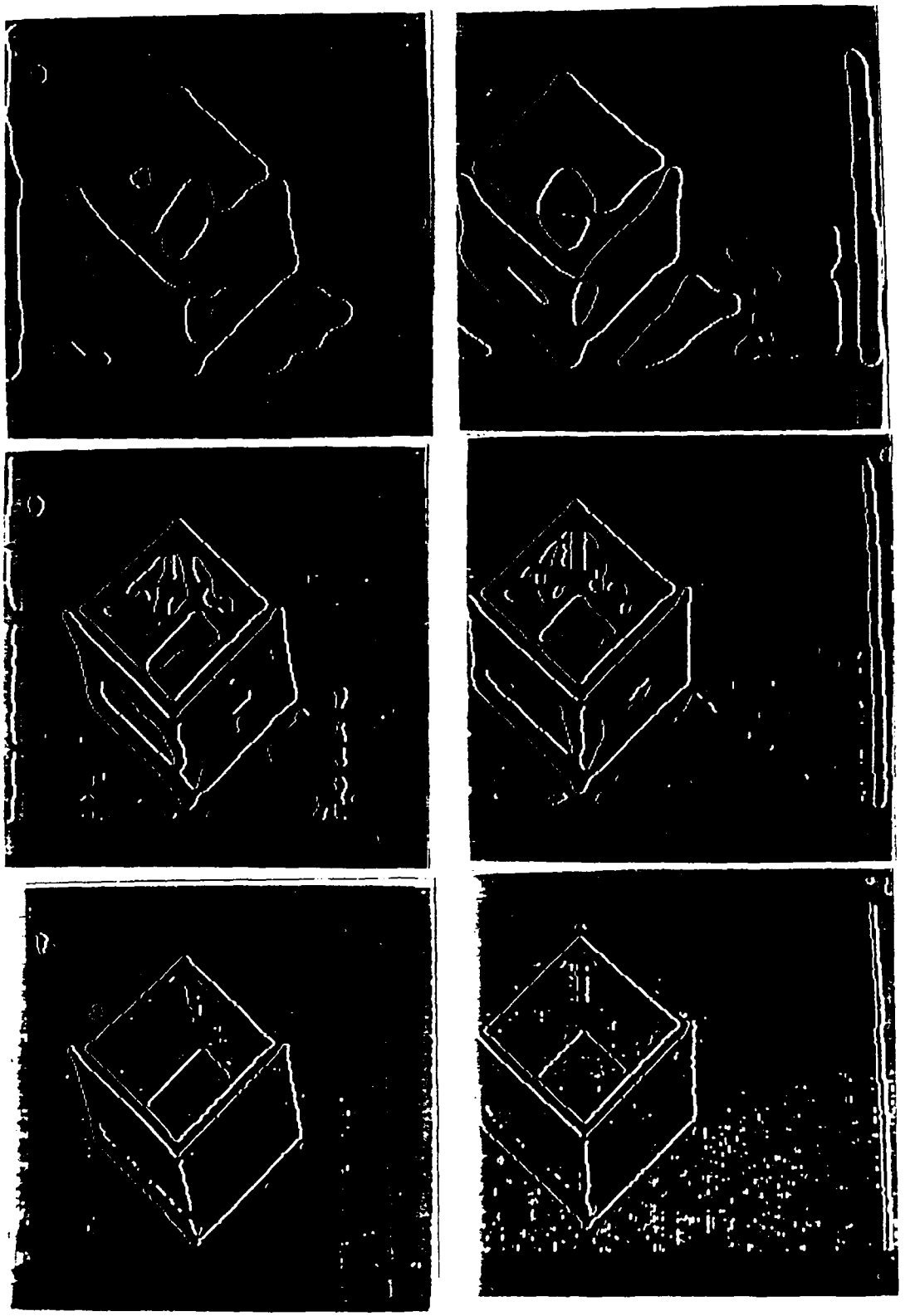


Fig. 6 Zero-Crossings of the box stereo image, for σ = 1.5, 3, and 6, a threshold of T=20 [60].

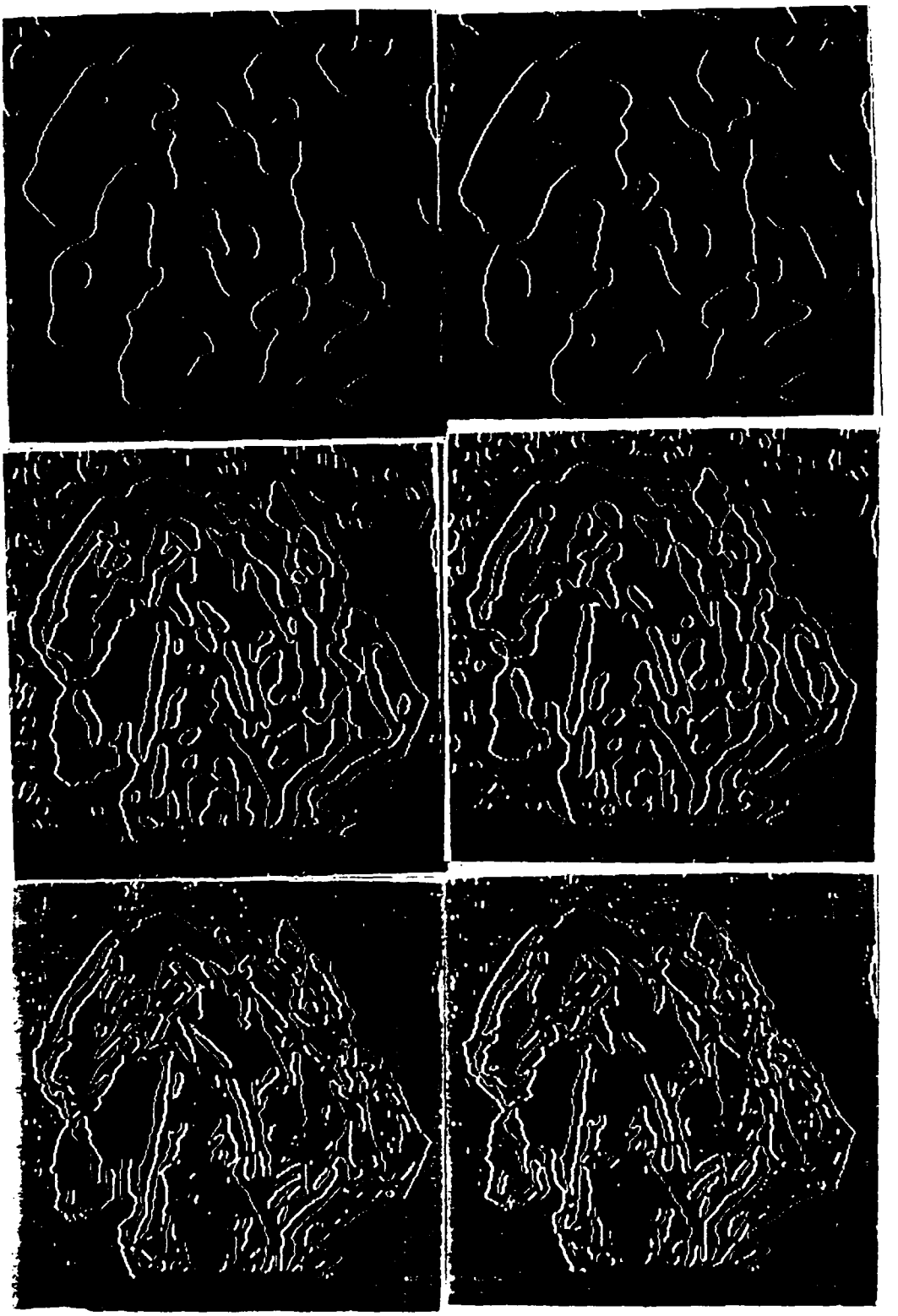


Fig. 7 Zero-Crossings of the paper-wad stereo image, for σ = 1.5, 3, and 6, a threshold of T=20 [60].

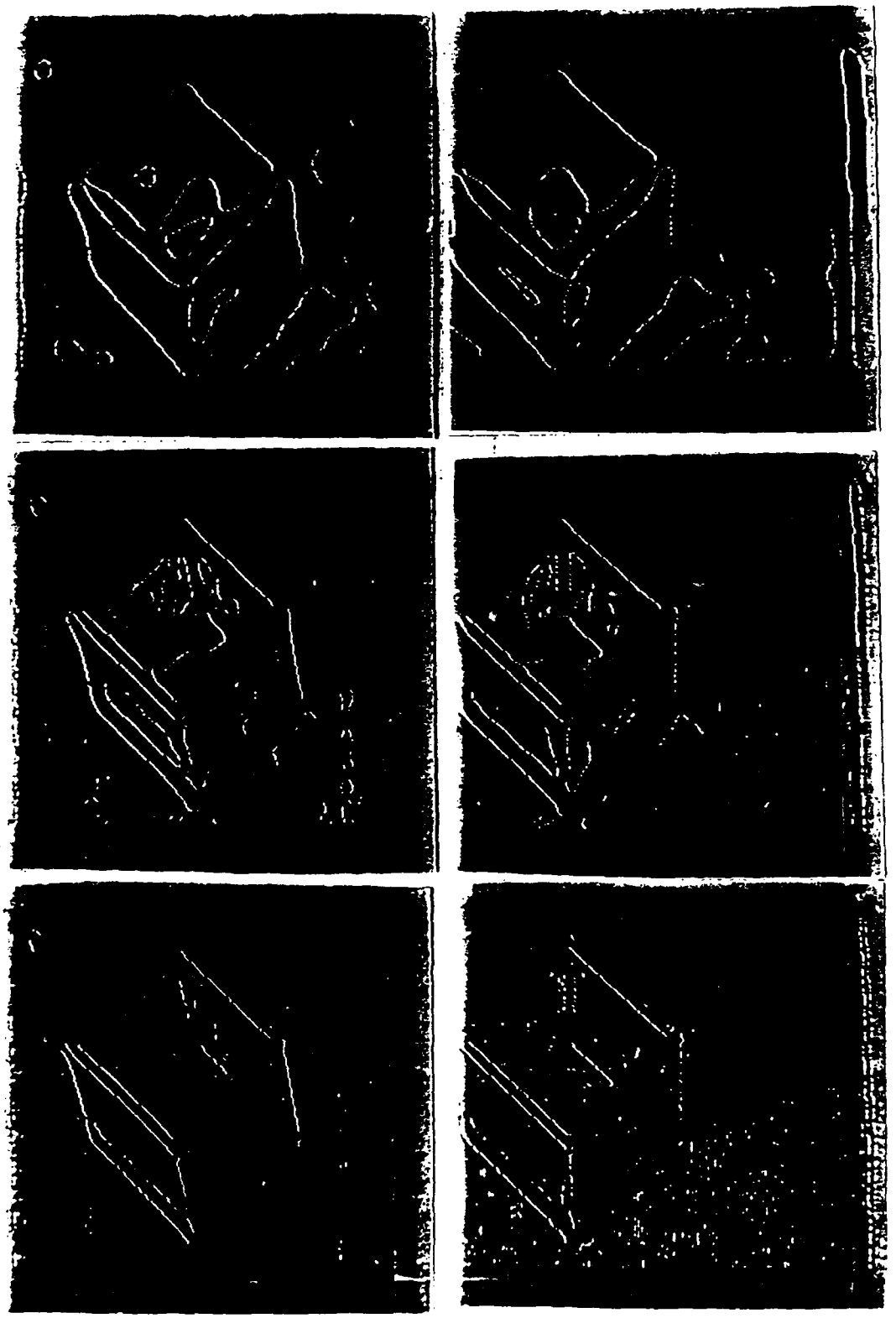


Fig. 8 Gray level coded representation of the zero-crossing's orientation of the box stereo image, for σ = 1.5, 3, and 6 [60].

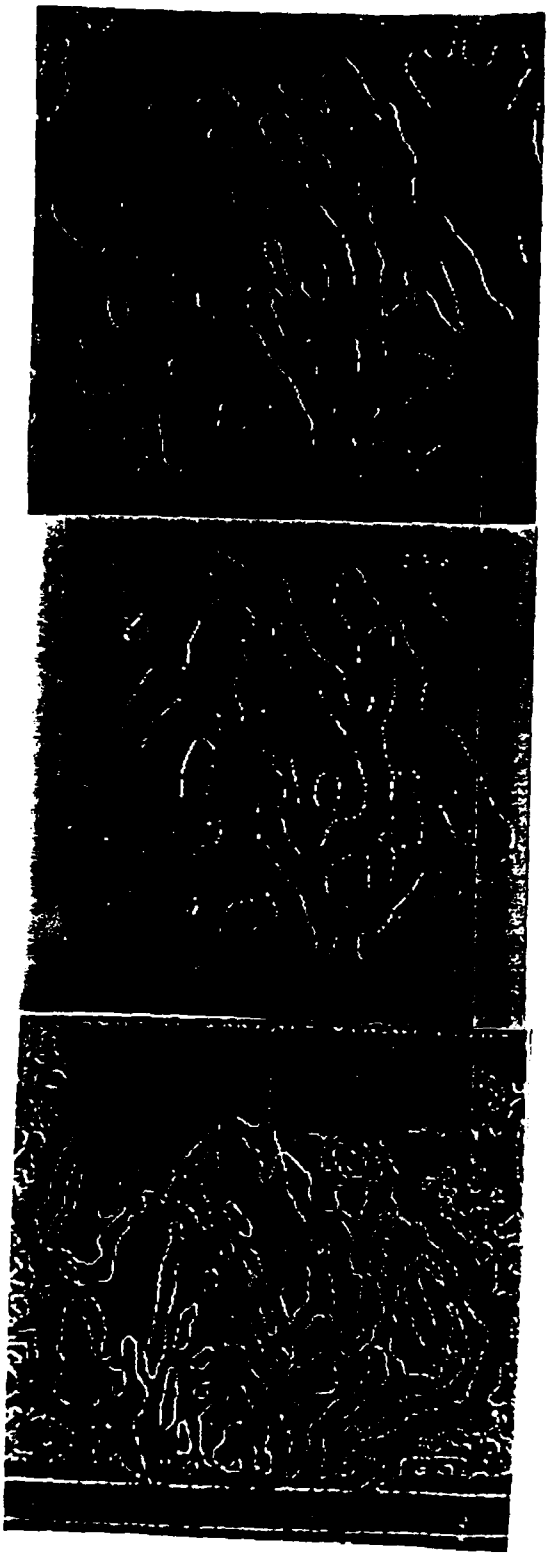


Fig. 9 Gray level coded representation of the zero-crossing's orientation of the paper-wad stereo image, for σ = 1.5, 3, and 6 [60].

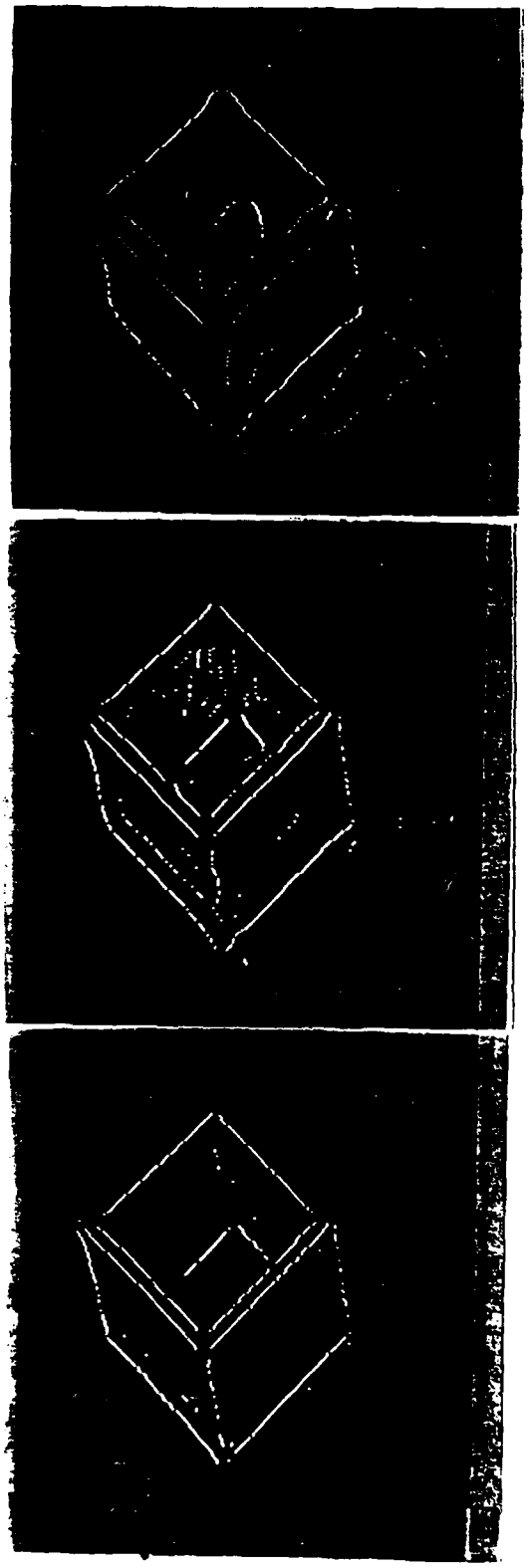


Fig. 10 Shows disparity values of the box image for each channel, disparities are displayed as gray level [60].

Fig. 11 Shows disparity values of the paper-wad image for each channel, disparities are displayed as gray level [60].

APPENDIX III

Vertebrate Binacular Vision System

In this appendix we study the human visual system in order to unravel the mystery behind perception of objects. There are a large number of unsolved questions, especially the mechanism and synaptic contacts (connection) between neurons in the visual cortex. However, in recent years advances have been made on the understanding of visual pathways and identifying specific cells in the visual cortex [68] - [72]. In this appendix we review the mechanism of the human visual system and the experimental results, based upon neurological findings, about the visual cortex of monkey and the cat.

1. Primitive Retina

The vertebrate retina consists of five distinct neuron cells: rod and cone photoreceptors, horizontal cells, bipolar cells, amacrine cells, and ganglion cells. Each of these cells perform a specific task. Figure 1 is a simplified illustration of the eye and the retina. The human retina contains two types of photoreceptors, rods and cones. Cones detect form and color, and are responsible for day vision. Rods mediate night vision and respond to stimuli that are too weak to excite cone receptors. These photoreceptors are distributed over the retina non-uniformly. For example, at the fovea of the retina, which defines the visual axis of the eye, the receptors are mainly cones and are packed very close by to each other. They are responsible for highly detailed and exact vision. At the periphery there are many more rods than cones and distribution is not as packed as in the fovea.

Figure 2 shows the synaptic contact between the photoreceptors and the next level of neurons. Both rods and cones make direct synaptic contact with a class of interneurons called the bipolar cells. There are two types of cone bipolar

cells, on-center and off-center, each of which responds differently to the same transmitter response by a single cone. The on-center bipolar cell is depolarized (excited) by direct illumination of the cone, and the off-center bipolar cell is hyperpolarized (inhibited) by direct illumination of the same cone. The rods are connected to the rod bipolar cells and are not directly connected to the ganglion cells, but they make indirect connection with the help of amacrine interneurons.

The on-center and the off-center bipolar cells are directly connected to the corresponding on-center and off-center ganglion cells. There are three distinct ganglion cells present in the retina, these are known as X, Y, and W cells. The X cells have medium-sized cell bodies and small dendritic fields (postsynaptic region of a neuron conducting impulses), and participate in high-acuity vision. The Y cells, have the largest cell bodies, a large dendritic fields, and rapidly conducting axons (the process of a neuron conducting impulses). The Y cells respond only to large targets and are important in the initial analysis of crude form. The W cells have small cell bodies and large dendritic fields; these cells project to the superior colliculus and are involved in head and eye movements. The ganglion cell do not convey information about absolute level of illumination, but rather, they measure differences within their receptive fields by comparing the degree of illumination between the center and the surround (contrast).

In summary, there are two independent channels: the on-center and the off-center. Each of these in turn is sub-divided into X and Y channels. The Y channel responds transiently and only to large targets, particularly moving ones. The X channel responds to small targets and is involved in the detailed high-resolution analysis of the visual image. These channels are believed to be responsible for the coarse-to-fine fusion of a binacular pair of images. The receptive field of a cell is defined as the area of the periphery on the retina whose stimulation influences the firing of a neuron. The characteristics of these receptive fields are discussed in the following sections.

2. The Visual Pathway

Figure 3 shows the projection of the visual world onto each retina. Each retina is represented by two halfs: the nasal hemiretina and temporal hemiretina. The left half of the visual field projects on the nasal hemiretina of the left eye and on the temporal hemiretina of the right eye, and a similar process happens for the right half of the visual field. The overlapping visual field that is perceived by both eyes is known as the binocular zone of the visual field. These binocular views are combined by the brain to provide us with the ability of stereoscopic depth perception. At the periphery there is no overlap and only monocular vision is possible.

Figure 4 shows a simplified block diagram of the visual pathway. The right visual hemifield is projected onto the temporal hemiretina of the left eye and the nasal hemiretina of the right eye. The fibers from nasal hemiretina crosses to the opposite side at the optic chiasm and make connections with the left lateral geniculate nucleus. Thus, the left optic tract contains a complete representation of the right hemifield of vision and the information about the left visual hemifield is conveyed by the right optic tract.

As shown in Figure 4, about 20% to 30% of the fibers in the optic nerve connect to the superior colliculus (for eye and head movement controls). However, a large number of them are connected to the lateral geniculate nucleus; also fibers from other parts of the central nerous system converge to it.

In primates, the lateral geniculate nucleus consists of six layers of neurons separated by interviewing layers of axons and dendrites. As shown in Figure 5 the layers are numbered from 6 most dorsally to 1 most ventrally. Each layer receives input from one eye only: fibers from the contralateral nasal retina contact layers 6, 4, and 1; fibers from the ipsilateral temporal retina contact layers 5, 3, and 2. Thus there are six maps of the contralateral visual hemifield in vertical register. The cells in the lateral geniculate nucleus are very similar to the retinal ganglion cells, with concentric receptive fields, with cell characteristics of on-center and off-center, and with X and Y cells properties. The only major

difference between cells in the lateral geniculate nucleus and those in the retina is that the antagonisms between the surroundings and the center are slightly enhanced in the geniculate cells. The axons from the geniculate cells make synoptic contacts with the cells in the primary visual cortex.

3. Primary Visual Cortex

From the lateral geniculate nucleus, neurons project via the optic radiation to the primary visual cortex. Figure 6 shows a schematic diagrams of the visual projections from the retina to the various visual areas of the cerebral cortex. Experimental results are available on area 17 of the visual cortex of the cat and the monkey which are discussed in this section. The cells in the retina and the lateral geniculate nucleus have two distinct receptive fields: on-center and off-center. Experimental results indicate that both types of neurons respond optimally to light contrast. Kuffler [71] found that the receptive fields of the retinal ganglion cells are roughly circular and vary in size across the retina. The on-center cells have receptive fields with a central excitatory zone and an inhibitory surrounding. Shining a spot of light on the center of the field causes an increase in the spontaneous firing of an on-center cell. In contrast, a light stimulus that encircles this central zone inhibits the cells' firing. Thus, the most effective excitatory stimulus for this cell is a spot of light on the center of its receptive field, and the most effective inhibitory stimulus is a ring of light on the surrounds of the receptive field. The off-center cells have inhibitory center and excitatory surround, and their response to illumination is shown in Figure 7.

The fibers from lateral geniculate nucleus make synaptic contacts with the cells in area IV of the visual cortex. The layers IV is divided into three subregions (a,b,c). Most of the fibers from the lateral geniculate nucleus terminate at the IVc. The cells in layer IVc are very similar to the cells encountered in the lateral geniculate and the retina. However, Hubel and Wiesel [69] have identified cortical cells in area 17 that lie above or below layer IVc that have receptive fields that are stimulated by directional line or bar of light. They found that

there are several cells known as simple and complex cells with receptive fields that have distinct characteristics. For example, Figure 8 shows a simple cell, the best stimulus for this cell is a vertically oriented light bar in the center of its receptive field. Other stimulus with different orientation are less effective or ineffective in exciting this cell. The complex cells are usually larger than those of simple cells but also have axis of orientation. The position of the stimulus within the receptive field is not crucial because there are no clearly defined excitatory or inhibitory zones. There is another complex cell which is better known as hypercomplex cell which are prominent in area 18 of the visual cortex but can also be seen in area 17. These hypercomplex cells respond best to stimulus such as corners or a line that stops.

Hubel and Wiesel have also discovered that the primary visual cortex, like the somatic sensory cortex, is organized into narrow columns Figure 9. Each column contains cells in layers IVc with concentric receptive fields. Above and below there are many simple cells and complex cells with almost identical retinal positions and identical axis of orientation. Detailed mapping of sets of adjacent columns by Hubel and Wiesel, using tangential penetrations with microelectrodes has revealed a very precise organization with an orderly shift in axis of orientation of about 10 degrees from one column to the next. There are also a number of other columns with different functional properties.

4. Binocular Interaction between Neurons

In the visual cortex the input from the two eyes are combined into one image, the exact processes is not completely known. However, it is believed that the fusion is performed by binocular complex cells. The cells in the retina, lateral geniculate nucleus, and a majority of simple cells in layer IVc are monocular, that is, they receive stimulation from exactly one of the two eyes. But about half of the complex cells in deeper layers of visual cortex are binocular, in that they can be influenced independently by both eyes and are beleived to be responsible for fusion of binacular images see Figure 10.

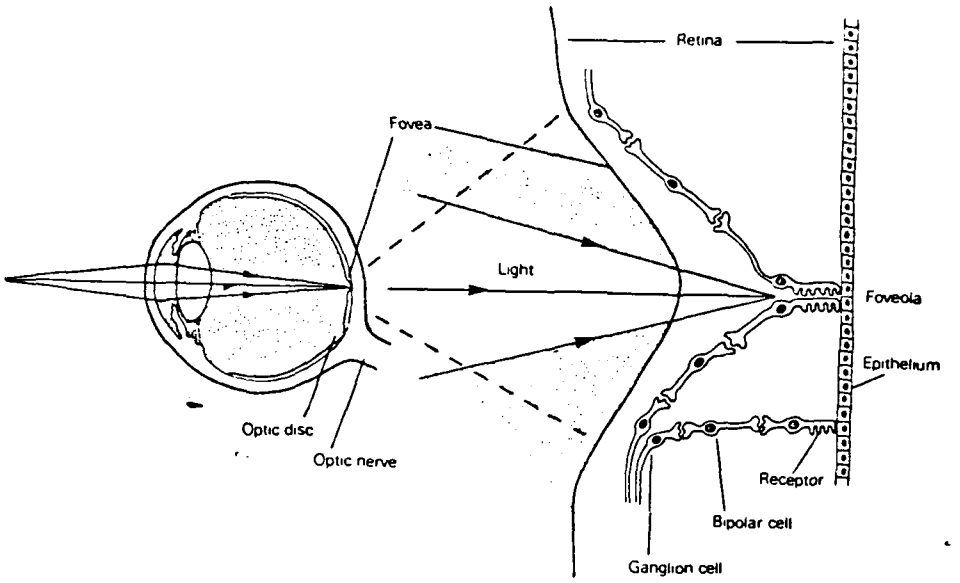


Fig. 1 A simplified illustration of the eye and the retina [72].

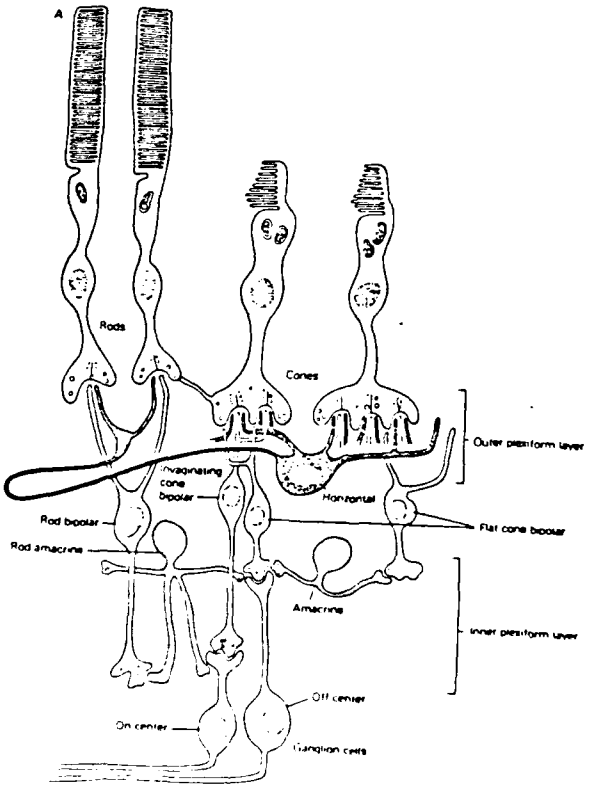


Fig. 2 A hierarchical synaptic connections of the cells in the retina [70].

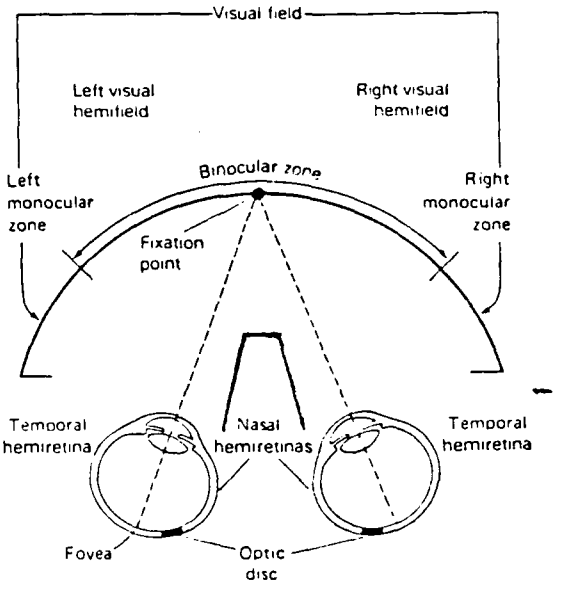


Fig. 3 Projection of the visual field onto each retina [72].

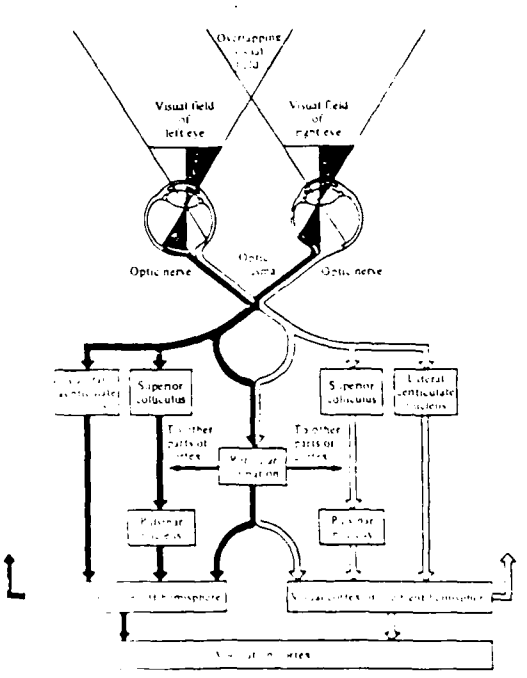


Fig. 4 Major known interconnections in the visual pathway [70].

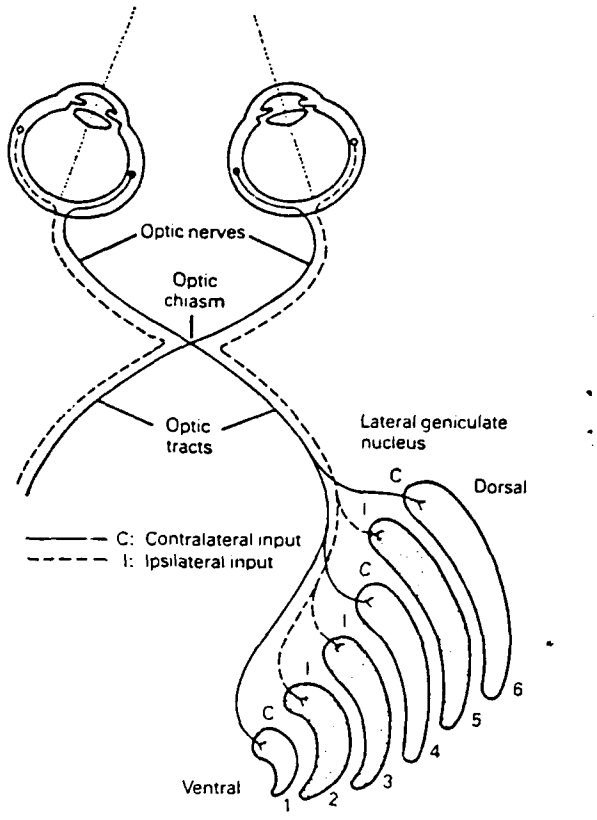


Fig. 5 Representation of the lateral geniculate nucleus and its connection with optic nerves [72].

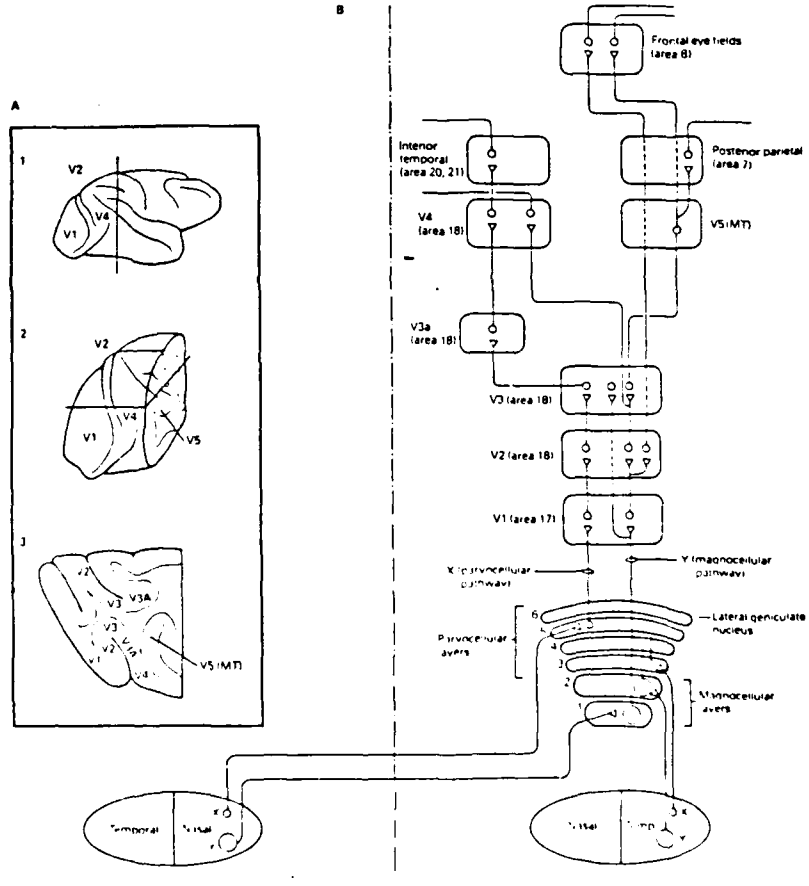


Fig. 6 Projection of the retina to the various visual areas of the cerebral cortex [72].

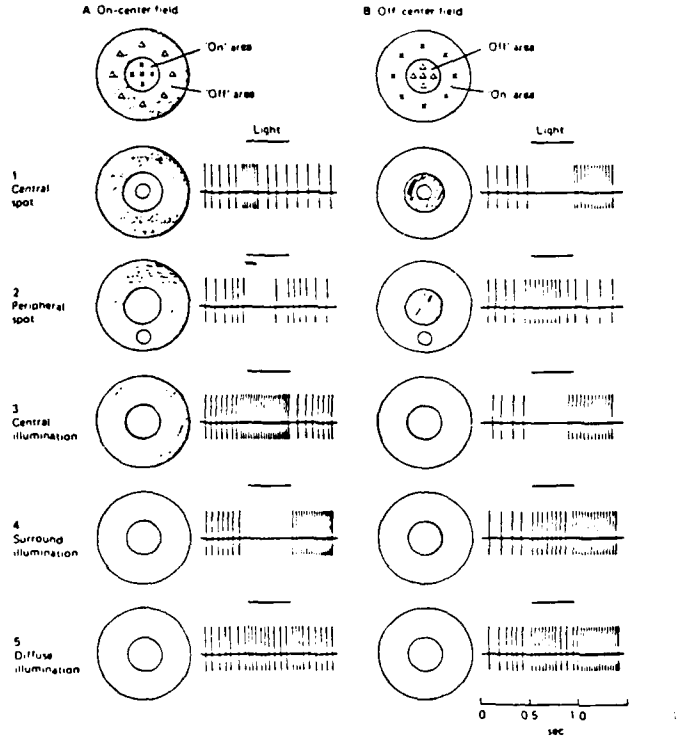


Fig. 7 The on and off-center receptive fields of the cells in the retina and the lateral geniculate nucleus [71].

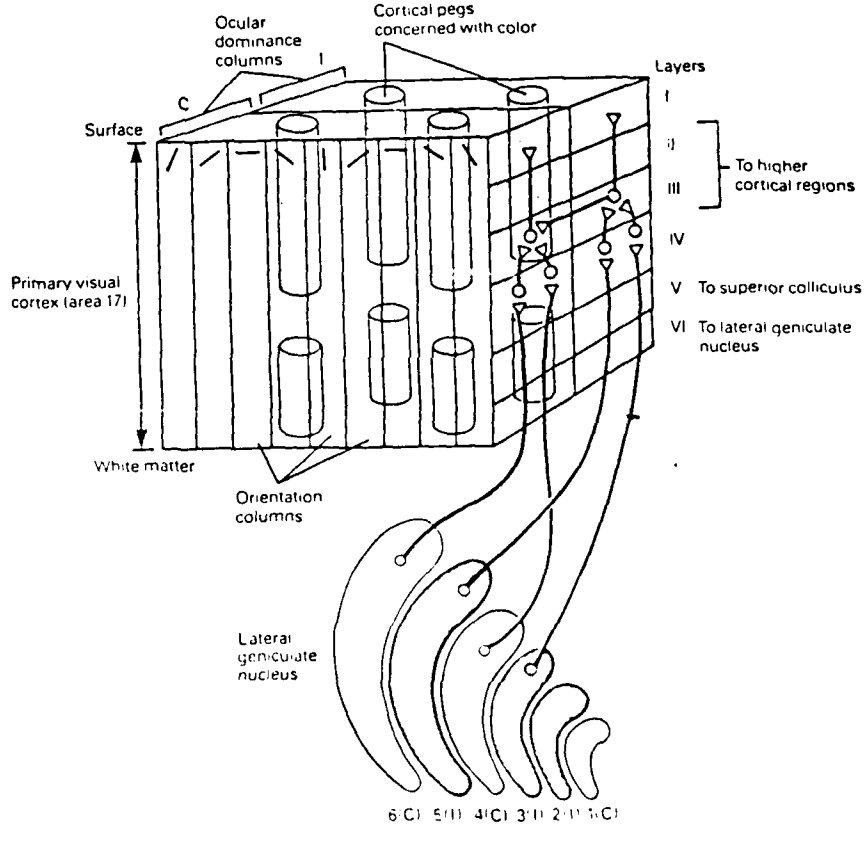


Fig. 9 The primary visual cortex is organized into a set of orientation columns and a set of ocular dominance columns [69].

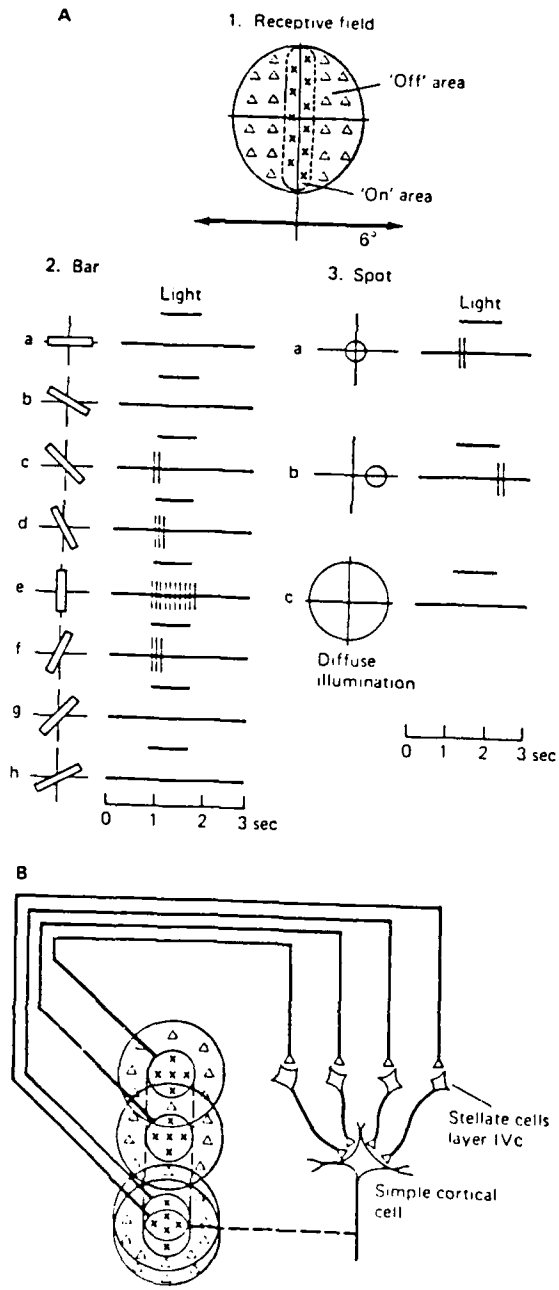


Fig. 8 The receptive field of a simple cell in the primary visual cortex [69].

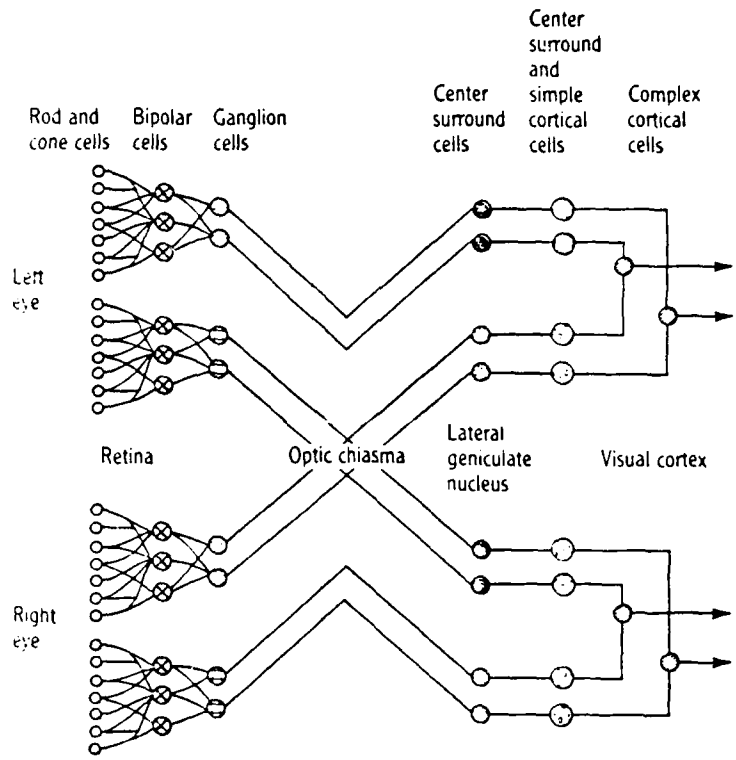


Fig. 10 The path from receptor cells to the visual cortex in the human stereo system [76].

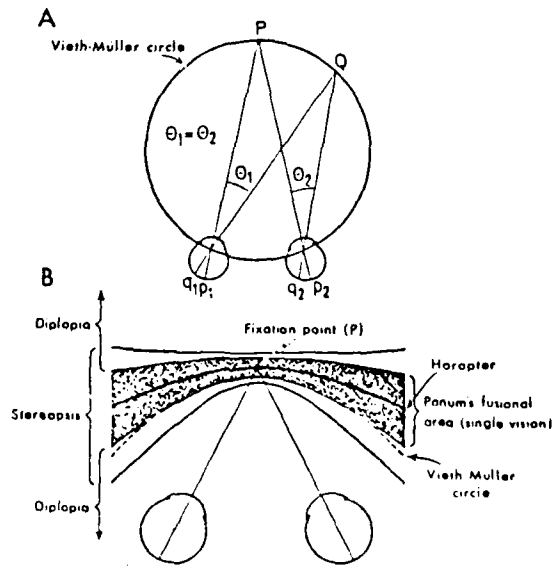


Fig. 11 Illustration of the Horopter, the Vieth-Muller Circle and the Panum's Fusional area [72].

APPENDIX IV

Stereoscopic Photogrammetry

Photogrammetry is defined as a technology for obtaining and interpreting photographic images [66]. Photographs are mostly aerial (taken from an airborne vehicle) or terrestrial photos (taken from earth-based cameras). Aerial photography is commonly classified as either vertical or oblique. Vertical photos are taken with the camera axis directed as nearly vertical as possible. Figure 1 shows a vertical aerial photograph. Oblique aerial photographs are exposed with the camera axis intentionally tilted away from vertical. Figure 2 is an example of a high oblique photograph.

Vertical aerial photographs are usually taken along a series of parallel airflight passes called flight strips. The photographs are normally exposed in such a way that each successive photograph overlaps by about 50 to 65 percent of the previous photo. This lapping along the flight strip is called end lap as shown in Figure 3. The pair of photos is called a stereopair, an example of which is shown in Figure 4. Stereo images are extremely important to photogrammetist because 3-D information about the topography can be accurately measured.

A large number of instruments, such as terrestrial stereo cameras, aerial cameras, stereoscopes and stereo plotters have been developed by photogrammetrists to obtain stereo photos and to interpret them. In this appendix we are only interested in aerial stereoscopic images.

1. Cameras

The essential requirement of any photogrammetric aerial camera is a lens of high geometric quality. They must be capable of exposing in rapid succession a great number of photographs to exact specifications. Aerial cameras may be

categorized as single-lens frame cameras, multi-lens frame cameras, strip cameras (used for continuous photography of a strip of terrain), and panoramic cameras. Figure 5 is an example of a two single-lens frame cameras mounted together and operated simultaneously. A pair of stereo images can be obtained by this camera. Terrestrial cameras are employed to obtain stereo images usually in special situations such as deep gorges or rugged mountains that are difficult to map from aerial photography. Figure 6 shows a terrestrial stereometric camera with 120 cm fixed baseline.

2. Stereoscopes and their Applications

It is very difficult to view photographs stereoscopically without the aid of an optical device. An instrument called stereoscope can be used to perceive a pair of stereo images in 3-D stereo model. The simplest stereoscope consists of two convex lenses mounted on a frame as shown in Figure 7.

The spacing between the lenses can be varied to accommodate various eye bases. Stereoscopes are used to locate the conjugate point (same targets on the two images) and when the appropriate disparity is obtained, the X and Y and Z coordinates of the target can be calculated (see Appendix I for equations). Figure 8 illustrates a packet stereoscope.

To locate the conjugate points on a stereo model, two small identical marks etched on clear glass called half marks are placed over the photographs, one on the left photo and other on the right photo, as illustrated in Figure 9. Viewing through a stereoscope the spacing of the half marks (parallex of the half marks) is varied by moving one of the halfmarks until a single floating mark appears to rest exactly on the terrain. By noting the locations of the half marks the disparity can be calculated and then the actual coordinates can be obtained.

3. Stereoplotters

Stereoscopic plotting instruments have been designed for a number of years in order to be able to obtain accurate object point positions from their corresponding image positions in a stereo pair. The basic concept for stereoplotter is that transparencies (diapositives) obtained from a pair of stereo images are placed in the two stereoplotter projectors as shown in Figure 10. Light rays are projected through them, and when rays from corresponding images on the left and right diapositives intersect below, they create a stereomodel. Some adjustment has to be performed in order to bring the pair of stereo images into correspondence.

A stereoplotter is said to consist of three major components: a projection system, a viewing system and a measuring (or tracing) system, which enables measurements of the stereomodel to be made and recorded. The most important part of the stereoplotter is the tracing system. To calculate the position of any point in the stereomodel, a reference mark in the center of a "platen" (white disk mounted on the tracing table) is brought into coincidence with model points. This is done by adjusting the X, Y, and Z coordinates of the platen until the reference mark appears to rest exactly on the desired point in the stereomodel. A pencil point which is vertically beneath the reference mark is then lowered to record the planimetric position of the point on the map. In some stereoplotters the X, Y, and Z coordinates are recorded automatically but adjustment of platen is done by the operator. In more complicated automated stereoplotters, image correlators are used to perform stereoscopic measurements with no human operator adjusting the floating mark. The image correlator is used to obtain the corresponding conjugate points on the stereo images. Research is needed to develop more accurate stereo vision techniques than just a simple image correlator which at some image points will perform very poorly.

4. X-Ray Stereo Photogrammetry

X-ray machines have been used for a number of applications such as in medical professions and industrial applications. Objects may be viewed stereoscopically from stereopairs of radiographs [67]. The exact 3-D location of foreign objects in the body such as bullets, pins, or tracks can be located from stereo radiographs.

Figure 11 shows a stereoradiograph of human skull and Figure 12 represents a side view of the geometry of an X-ray stereoradiograph. In industry, X-rays have a host of applications including examination of radioactive fuels, castings, and locating pipes and wires in buildings.

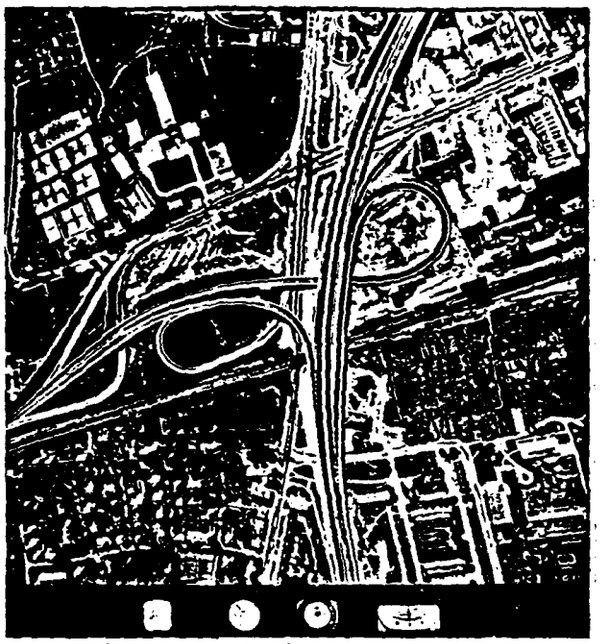


Fig. 1 Shows a vertical aerial photograph [66 PP. 6].

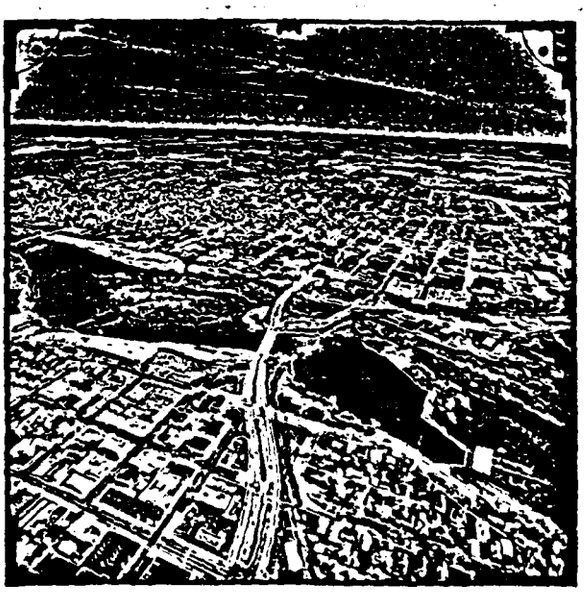


Fig. 2 Shows a high oblique aerial photograph [66 PP. 9].

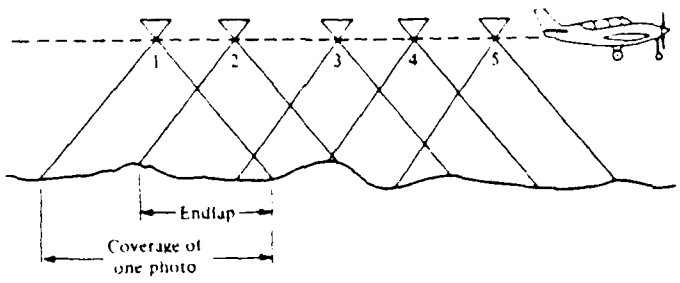


Fig. 3 End lap of photographs in a flight strip [66 PP. 10].

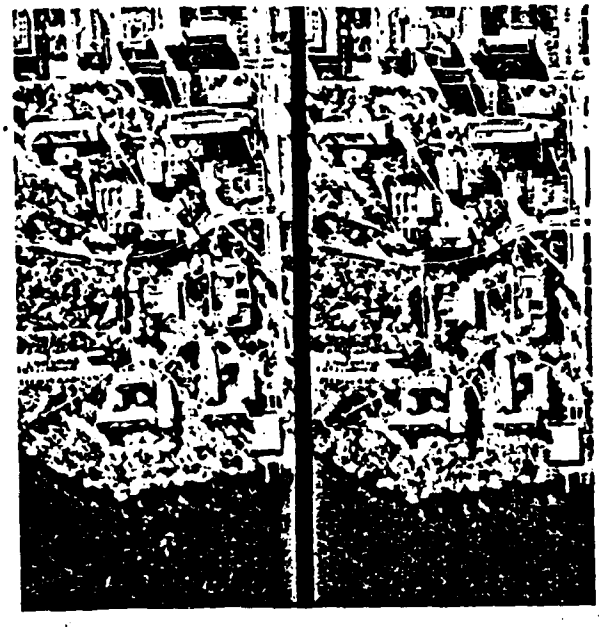


Fig. 4 A pair of aerial stereo images [66 PP. 516].

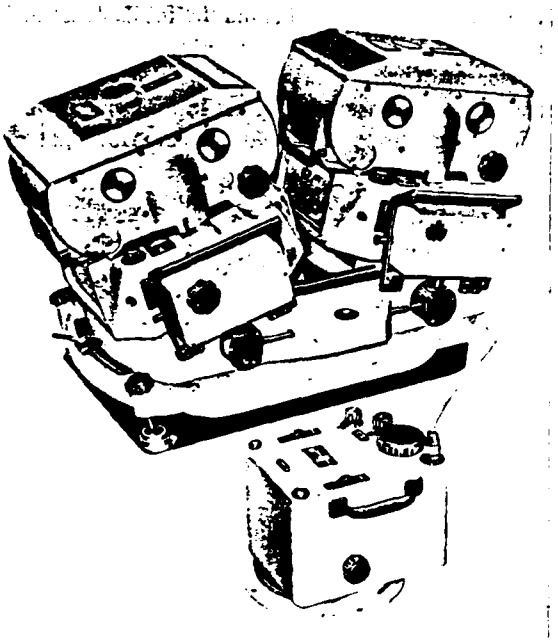


Fig. 5 A convergent aerial camera to obtain stereo images [66 PP. 67].

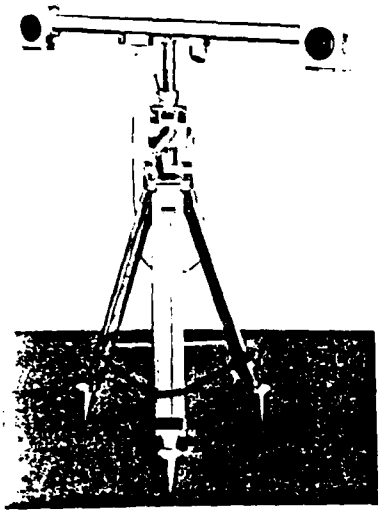


Fig. 6 A terrestrial stereometric camera [66 PP. 483].

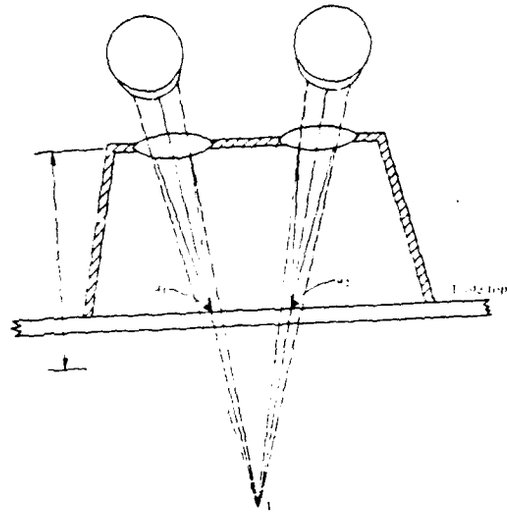


Fig. 7 Schematic diagram of a pocket stereoscope [66 PP. 148].

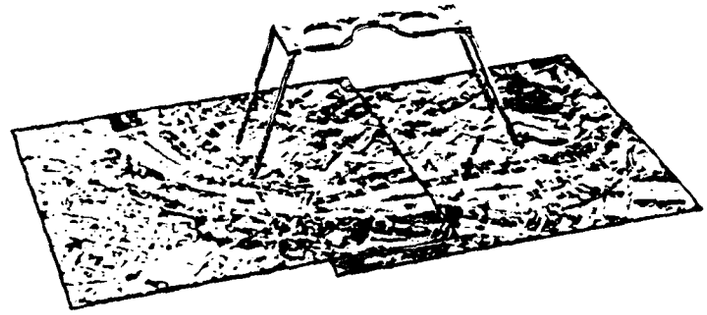


Fig. 8 A pocket stereoscope [66 PP. 147].

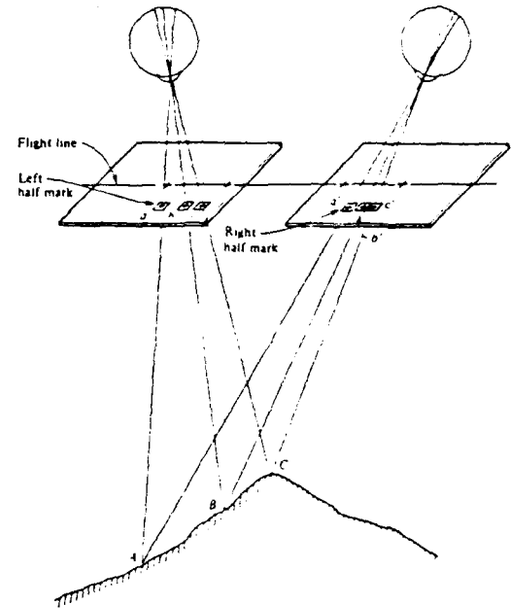


Fig. 9 The principle of the floating mark [66 PP. 165].

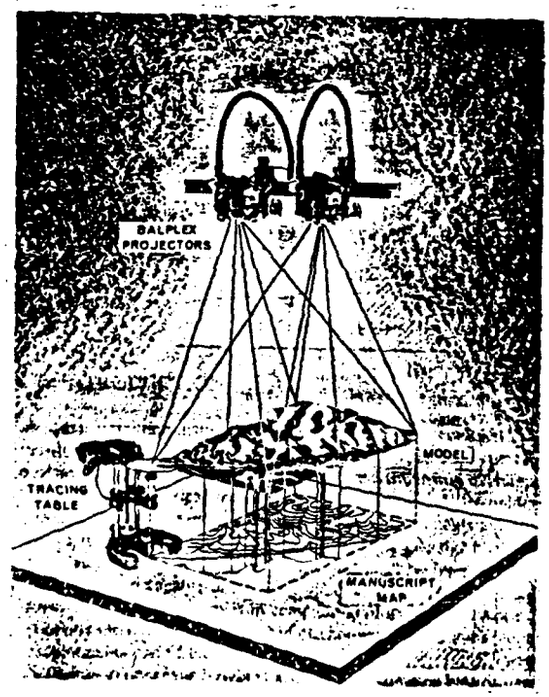


Fig. 10 The basic concept of a stereoplotter [66 PP. 266].

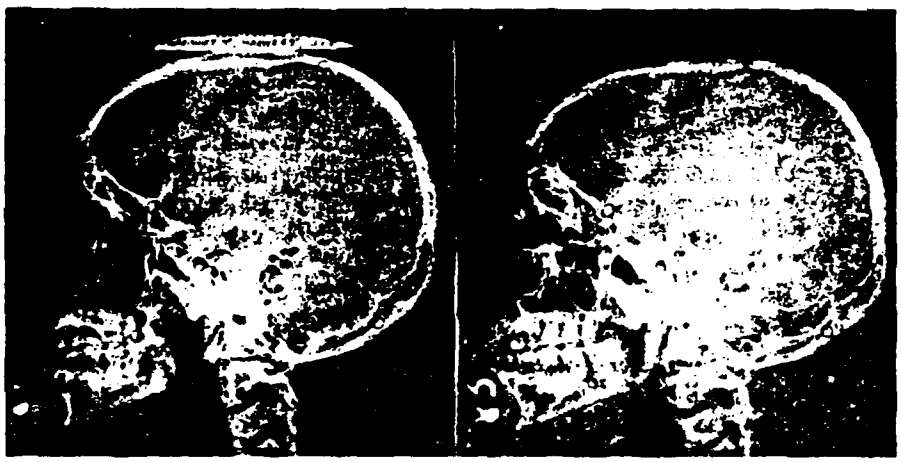


Fig. 11 A stereoradiograph of a human skull [66 PP. 504].

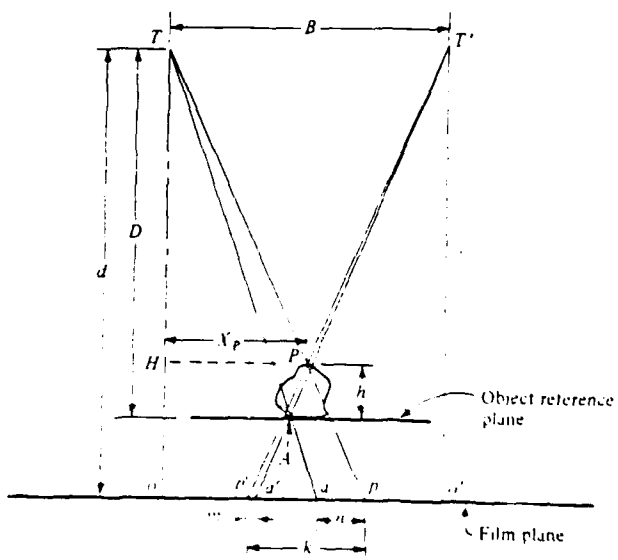


Fig. 12 A side view of the geometry of an x-ray stereoradiograph [66 PP. 505].